APPLICATION

FOR

U.S. LETTERS PATENT

DEVELOPMENT OF NOVEL ANTI-MICROBIAL AGENTS BASED ON BACTERIOPHAGE GENOMICS

CERTIFICATE OF MAILING
(37 C.F.R. §1.10)

I hereby certify that this paper (along with any referred to as being attached or enclosed) is being deposited with the United States Postal Service on the date shown below with sufficient postage as 'Express Mail Post Office To Addressee' in an envelope addressed to the Assistant Commissioner for Patents, Washington, D.C. 20231.

EL356074807US
Express Mail Label No.

Wesley B. Ames
Name of Person Mailing Paper

December 2, 1999
Date of Deposit

W.B. Ames
Signature of Person Mailing Paper

DESCRIPTION

Development of Novel Anti-Microbial Agents Based on Bacteriophage Genomics

RELATED APPLICATIONS

5   This application is a continuation-in-part of U.S. Application No. 09/407,804, filed September 28, 1999, entitled DNA SEQUENCES FROM STAPYLOCOCCUS AUREUS BACTERIOPHAGE 77 THAT ENCODE ANTI-MICROBIAL POLYPEPTIDES, and claims the benefit of U.S. Provisional Application No. 60/110,992, filed December 3, 1998, entitled DEVELOPMENT OF NOVEL
10   ANTIMICROBIAL AGENTS BASED ON BACTERIOPHAGE GENOMICS, which are hereby incorporated by reference in their entireties, including drawings.

BACKGROUND OF THE INVENTION

15   The present invention relates to the field of antibacterial agents and the treatment of infections of animals or other complex organisms by bacteria.

          The frequency and spectrum of antibiotic-resistant infections have, in recent years, increased in both the hospital and community. Certain infections have become
20   essentially untreatable and are growing to epidemic proportions in the developing world as well as in institutional settings in the developed world. The staggering spread of antibiotic resistance in pathogenic bacteria has been attributed to microbial genetic characteristics, widespread use of antibiotic drugs, and changes in society that enhance the transmission of drug-resistant organisms. This spread of drug resistant
25   microbes is leading to ever increasing morbidity, mortality and health-care costs.
          Ironically, it is the very success of antibiotics, resulting in their widespread use, that has contributed the most to rising numbers of drug resistant bacterial strains. The longer a bacterial strain is exposed to a drug, the more likely it is to acquire resistance. Today, a total of 160 antibiotics, all based on a few basic chemical
30   structures and targeting a small number of metabolic pathways, have found their way to market. Over-prescription of these drugs, as well as the failure of patients to comply with the complete antibiotic regimen, has lead to the rapid emergence of antibiotic resistant strains. Such misuse of prescriptions, careless use of antibiotics in virtually all commercial production of beef and fowl, and changing societal
35   conditions, such as the growth of day-care centers, increased long-term care in hospitals, and increased mobility of the population, has provided an environment

SD-63311.4

where drug-resistant microbes can emerge and spread. Thus, virtually all common infectious bacteria are becoming, or have already become, resistant to one or more groups of antibiotics. Such resistance now reaches all classes of antibiotics currently in use, including: β-lactams, fluoroquinolones, aminoglycosides, macrolide peptides,

5   chloramphenicol, tetracyclines, rifampicin, folate inhibitors, glycopeptides, and mupirocin.

Over the last 45 years bacteria have adapted genetically to avoid the destruction/alteration of the essential pathways that these chemotherapeutic agents target. Antibiotic resistant bacterial strains are now emerging at a higher rate than the

10   rate at which new antibiotics are being developed. The consequence of this dilemma has been a dramatic increase in the cost of treating infections what would otherwise easily succumb to routine antibiotic therapy. Furthermore, and perhaps most importantly, the emergence of multiple drug resistant pathogenic bacteria has led to a significant increase in morbidity and mortality, particularly in institutional settings.

15   Most major pharmaceutical companies have on-going drug discovery programs for novel anti-microbials. These are based on screens for small molecule inhibitors (natural products, bacterial culture media, libraries of small molecules, combinatorial chemistry) of crucial metabolic pathways of the micro-organism of interest (e.g., bacteria, fungi, parasites, worms). The screening process is largely for

20   cytotoxic compounds and in most cases is not based on a known mechanism of action of the compounds. Pharmaceutical companies have large programs in this area. Classical drug screening programs are being exhausted and many of these pharmaceutical companies are looking towards rational drug design programs.

Several small to mid-size biotechnology companies as well as large

25   pharmaceutical companies have developed systematic high-throughput sequencing programs to decipher the genetic code of specific micro-organisms of interest. The goal is to identify, through sequencing, unique biochemical pathways or intermediates that are unique to the microorganism. Knowledge of this may, in turn, form the rationale for a drug discovery program based on the mechanism of action of the

30   identified enzymes/proteins. Genome Therapeutics Corp., The Institute for Genome Research, Human Genome Sciences Inc., and other companies have such sequencing programs in place. However, one of the most critical steps in this approach is the ascertainment that the identified proteins and biochemical pathways are 1) non-redundant and essential for bacterial survival, and 2) constitute suitable and accessible

35   targets for drug discovery.

SUMMARY OF THE INVENTION

While animals such as humans are, on occasion, infected by pathogenic bacteria, bacteria also have natural enemies. A number of host-specific viruses,
5   known as bacteriophages or phages, infect and kill bacteria in the natural environment. Such bacteriophages generally have small compact genomes and bacteria are their exclusive hosts. Many known bacteria are host to a large number of bacteriophages that have been described in the literature. During the 1940's - 1960's, phage biology was an area of active research. As a testimony to this, the study of
10   phages which infect and inhibit the enteric bacterium *Escherichia coli* (*E. coli*) contributed much to the early understanding of molecular biology and virology.

This invention utilizes the observation that bacteriophages successfully infect and inhibit or kill host bacteria, targeting a variety of normal host metabolic and physiological traits, some of which are shared by all bacteria, pathogenic and
15   nonpathogenic alike. The term "pathogenic" as used herein denotes a contribution to or implication in disease or a morbid state of an infected organism. The invention thus involves identifying and elucidating the molecular mechanisms by which phages interfere with host bacterial metabolism, an objective being to provide novel targets for drug design. Whether the phage blocks bacterial RNA transcription or translation,
20   or attacks other important metabolic pathways, such as cell wall assembly or membrane integrity, the basic blueprint for a phage's bacteria-inhibiting ability is encoded in its genome and can be unlocked using bioinformatics, functional genomics, and proteomics. By these means, the invention utilizes sequence information from the genomics of bacteriophage to identify novel antimicrobials that
25   can be further used to actively and/or prophylactically treat bacterial infection.

Two important components of the invention thus are: i) the identification of bacteria-inhibiting phage open reading frames ("ORF"s) and corresponding products that can be used to develop antibiotics based on amino acid sequence and secondary structural characteristics of the ORF products, and ii) the use of bacteriophages to map
30   out essential bacterial target genes and homologs, which can in turn lead to the development of suitable anti-microbial agents. These two avenues represent new and general methods for developing novel antimicrobials.

The invention thus concerns the identification of bacteriophage ORFs that supply bacteria-inhibiting functions. In this regard, use of the terms "inhibit",
35   "inhibition", "inhibitory", and "inhibitor" all refer to a function of reducing a biological activity or function. Such reduction in activity or function can, for example, be in connection with a cellular component, *e.g.*, an enzyme, or in

SD-63311.4

connection with a cellular process, *e.g.*, synthesis of a particular protein, or in connection with an overall process of a cell, *e.g.*, cell growth. In reference to bacterial cell growth, for example, an inhibitory effect (i.e., a bacteria-inhibiting effect) may be bacteriocidal (killing of bacterial cells) or bacteriostatic (i.e., stopping or at least

5   slowing bacterial cell growth). The latter slows or prevents cell growth such that fewer cells of the strain are produced relative to uninhibited cells over a given period of time. From a molecular standpoint, such inhibition may equate with a reduction in the level of, or elimination of, the transcription and/or translation of a specific bacterial target(s), or reduction or elimination of activity of a particular target

10   biomolecule.

It is particularly advantageous to evaluate a plurality of different phage ORFs for inhibitory activity which may be from one, but is preferably from a plurality of different phage. For example, evaluating ORFs from a number of different phage of the same bacterial host provides at least two advantages. One is that the multiple

15   phages will provide identification of a variety of different targets. Second, it is likely that multiple phage will utilize the same cellular target.

As used herein, the terms "bacteriophage" and "phage" are used interchangeably to refer to a virus which can infect a bacterial strain or a number of different bacterial strains.

20   In the context of this invention, the term "bacteriophage ORF" or ""phage ORF" or similar term refers to a nucleotide sequence in or from a bacteriophage. In connection with a particular ORF, the terms refer an open reading frame which has at least 95% sequence identity, preferably at least 97% sequence identity, more preferably at least 98% sequence identity with an ORF from the particular phage

25   identified herein (e.g., with an ORF as identified herein) or to a nucleic acid sequence which has the specified sequence identify percentage with such an ORF sequence.

A first aspect of the invention thus provides a method for identifying a bacteriophage nucleic acid coding region encoding a product active on an essential bacterial target by identifying a nucleic acid sequence encoding a gene product which

30   provides a bacteria-inhibiting function when the bacteriophage infects a host bacterium, preferably one that is an animal or plant pathogen, more preferably a bird or mammalian pathogen, and most preferably a human pathogen. The bacteriophage is an uncharacterized bacteriophage. Thus, the method excludes, for example, phage λ, φx174, m13 and other *E.coli*-specific bacteriophage that have been studied with

35   respect to gene number and/or function. It also excludes, for example, the nucleic acid coding regions described in Tables 13-14, and in preferred embodiments, excludes the phage in which those regions are naturally located. In preferred

embodiments of this and the other aspects of the present invention, the phage is *Staphylococcus aureus* phage 77, 3A, or 96.

In connection with bacteriophage, the term "uncharacterized" means that a certain bacteriophage's genome has not yet been fully identified such that the genes having function involved in inhibiting host cells have not been identified. In particular, phage for which the description of genomic or protein sequence was first provided herein are uncharacterized. Phage sequences for which host bacteria-inhibiting functions have been identified prior to the filing of the present application (or alternatively prior to the present invention) are specifically excluded from the aspects involving utilization of sequences from uncharacterized bacteriophage, except that aspects may involve a plurality of phage where one or more of those phage are uncharacterized and one or more others have been characterized to some extent. A number of different bacteria-inhibiting phage ORFs are indicated in Tables 12-14. The phage ORFs or sequences identified therein are not within the term "uncharacterized; alternatively, in preferred embodiments the phage containing those ORFs are excluded from this term. Further, any additional phage ORFs (or alternatively the phage which contain those ORFs) which have previously been described in the art as bacteria-inhibiting ORFs are expressly excluded; those ORFs or phage are known to those skilled in the art and the exclusion can be made express by specifically naming such ORFs or phage as needed (likewise for uncharacterized targets as described below). For the sake of brevity, such a listing is not expressly presented, as such information is readily available to those skilled in the art.

Stating that an agent or compound is "active on" a particular cellular target, such as the product of a particular gene, means that the target is an important part of a cellular pathway which includes that target and that the agent acts on that pathway. Thus, in some cases the agent may act on a component upstream or downstream of the stated target, including on a regulator of that pathway or a component of that pathway.

By "essential", in connection with a gene or gene product, is meant that the host cannot survive without, or is significantly growth compromised, in the absence depletion, or alteration of functional product. An "essential gene" is thus one that encodes a product that is beneficial, or preferably necessary, for cellular growth *in vitro* in a medium appropriate for growth of a strain having a wild-type allele corresponding to the particular gene in question. Therefore, if an essential gene is inactivated or inhibited, that cell will grow significantly more slowly, preferably less than 20%, more preferably less than 10%, most preferably less than 5% of the growth rate of the uninhibited wild-type, or not at all, in the growth medium. Preferably, in

the absence of activity provided by a product of the gene, the cell will not grow at all or will be non-viable, at least under culture conditions similar to the *in vivo* conditions normally encountered by the bacterial cell during an infection. For example, absence of the biological activity of certain enzymes involved in bacterial cell wall synthesis

5   can result in the lysis of cells under normal osmotic conditions, even though protoplasts can be maintained under controlled osmotic conditions. In the context of the invention, essential genes are generally the preferred targets of antimicrobial agents. Essential genes can encode target molecules directly or can encode a product involved in the production, modification, or maintenance of a target molecule.

10   A "target" refers to a biomolecule that can be acted on by an exogenous agent, thereby modulating, preferably inhibiting, growth or viability of a cell. In most cases such a target will be a nucleic acid sequence or molecule, or a polypeptide or protein. However, other types of biomolecules can also be targets, *e.g.*, membrane lipids and cell wall structural components.

15   The term "bacterium" refers to a single bacterial strain, and includes a single cell, and a plurality or population of cells of that strain unless clearly indicated to the contrary. In reference to bacteria or bacteriophage, the term "strain" refers to bacteria or phage having a particular genetic content. The genetic content includes genomic content as well as recombinant vectors. Thus, for example, two otherwise identical

20   bacterial cells would represent different strains if each contained a vector, *e.g.*, a plasmid, with different phage ORF inserts.

Preferred embodiments involve expressing at least one recombinant phage ORF(s) in a bacterial host followed by inhibition analysis of that host. Inhibition following expression of the phage ORF is indicative that the product of the ORF is

25   active on an essential bacterial target. Such evaluation can be carried out in a variety of different formats, such as on a support matrix such as a solidified medium in a petri dish, or in liquid culture. Preferably a plurality of phage ORFs are expressed in at least one bacterium. The plurality of phage ORFs can be from one or a plurality of phage. With respect to a single phage or at least one phage in a plurality of phages,

30   the plurality of expressed ORFs preferably represents at least 10%, more preferably at least 20%, 40%, or 60%, still more preferably at least 80% or 90%, and most preferably at least 95% of the ORFs in the phage genome. Preferably, for a plurality of phage, the plurality of expressed ORFs preferably represents at least 10%, more preferably at least 20%, 40%, or 60%, still more preferably at least 80% or 90%, and

35   most preferably at least 95% of the ORFs in the phage genome of each phage. The plurality of phage ORFs can be expressed in a single bacterium, or in a plurality of

SD-63311.4

bacteria where one ORF is expressed in each bacterium, or in a plurality of bacteria where a plurality of ORFs are expressed in at least one or in all of the plurality of bacteria, or combinations of these.

5   In embodiments of the above aspect (as well as in other aspects herein) in which a plurality of phage are utilized, a plurality of phage have the same bacterial host species; have different bacterial host species; or both. The plurality of phage includes at least two different phage, preferably at least 3,4,5,6,8,10,15,20, or more different phage. Indeed, more preferably, the plurality of phage will include 50, 75, 100, or more phage. As described herein, the larger number of phage is useful to

10 provide additional target and target evaluation information useful in developing antibacterial agents, for example, by providing identification of a larger range of bacterial targets, and/or providing further indication of the suitability of a particular target (for example, utilization of a target by a number of different unrelated phage can suggest that the target is particularly stable and accessible and effective) and/or

15 can indicate alternate sites on a target which interact with different inhibitors.

Further embodiments involve confirmation of the inhibitor function of the phage ORF, such as by utilizing or incorporating a control(s) designed to confirm the inhibitory nature of the ORF(s) being evaluated. The control can, for example, be provided by expression of an inactive or partially inactive form of the ORF or ORF

20 product, and/or by the absence of expression of the ORF or ORF product in the same or a closely comparable bacterial strain as that used for expression of the test ORF. The reduced level of activity or the absence of active ORF product in the control will thus not provide the inhibition provided by a corresponding inhibitory ORF, or will provide a distinguishably lower level of inhibition. An inactivated or partially

25 inactivated control has a mutation(s), e.g., in the coding region or in flanking regulatory elements, that reduce(s) or eliminate(s) the normal function of the ORF. Thus, the inhibition of a bacterium following expression of a phage ORF is determined by comparison with the effects of expression of an inactivated ORF or the response of the bacteria in the absence of expression in the same or similar type

30 bacterium. Such determination of inhibition of the bacterium following expression of the ORF is indicative of a bacteria-inhibiting function. These manipulations are routinely understood and accomplished by those of skill in the art using standard techniques. In embodiments utilizing absence of expression of the ORF, the bacteria can, for example, contain an empty vector or a vector which allows expression of an

35 unrelated sequence which is preferably non-inhibitory. Alternatively, the bacteria may have no vector at all. Combinations of such controls or other controls may also be utilized as recognized by those skilled in the art.

SD-63311.4

In embodiments involving expression of a phage ORF in a bacterial strain, in preferred embodiments that expression is inducible. By "inducible" is meant that expression is absent or occurs at a low level until the occurrence of an appropriate environmental stimulus provides otherwise. For the present invention such induction

5   is preferably controlled by an artificial environmental change, such as by contacting a bacterial strain population with an inducing compound (*i.e.*, an inducer). However, induction could also occur, for example, in response to build-up of a compound produced by the bacteria in the bacterial culture, *e.g.*, in the medium. As uncontrolled or constitutive expression of inhibitory ORFs can severely compromise bacteria to the

10 point of eradication, such expression is therefore undesirable in many cases because it would prevent effective evaluation of the strain and inhibitor being studied. For example, such uncontrolled expression could prevent any growth of the strain following insertion of a recombinant ORF, thus preventing determination of effective transfection or transformation. A controlled or inducible expression is therefore

15 advantageous and is generally provided through the provision of suitable regulatory elements, *e.g.*, promoter/operator sequences that can be conveniently transcriptionally linked to a coding sequence to be evaluated. In most cases, the vector will also contain sequences suitable for efficient replication of the vector in the same or different host cells and/or sequences allowing selection of cells containing the vector,

20 *i.e.*, "selectable markers." Further, preferred vectors include convenient primer sequences flanking the cloning region from which PCR and/or sequencing may be performed.

As knowledge of the nucleotide sequence of phage ORFs is useful, *e.g.*, for assisting in the identification of phage proteins active against essential bacterial host

25 targets, preferred embodiments involve the sequencing of at least a portion of the phage genome in combination with the above methods. This can be done either before or after or independent of expression and inhibition of the ORF in the bacteria, and provides information on the nature and characteristics of the ORF. Such a portion is preferably at least 10%, 20%, 40%, 80%, 90%, or 100% of the phage genome. For

30 embodiments in which a plurality of phage are utilized, preferably each phage is sequenced to an extent as just specified.

Such sequencing is preferably accompanied by computer sequence analysis to define and evaluate ORF(s), ORF products, structural motifs or functional properties of ORF products, and/or their genetic control elements. Thus, certain embodiments

35 incorporate computer sequence analyses or nucleic acid and/or amino acid sequences. Further, existing data banks can provide phage sequence and product information which can be utilized for analysis and identification of ORFs in the sequence. Computer analysis may further employ known homologous sequences from other

SD-63311.4

species that suggest or indicate conserved underlying biochemical function(s) for the inhibitory or potentially inhibitory ORF sequence(s) being evaluated. This can include the sequences of signature motifs of identified classes of inhibitors.

In the context of the phage nucleic acid sequences, e.g., gene sequences, of this invention, the terms "homolog" and "homologous" denote nucleotide sequences from different bacteria or phage strains or species or from other types of organisms that have significantly related nucleotide sequences, and consequently significantly related encoded gene products, preferably having related function. Homologous gene sequences or coding sequences have at least 70% sequence identity (as defined by the maximal base match in a computer-generated alignment of two or more nucleic acid sequences) over at least one sequence window of 48 nucleotides, more preferably at least 80 or 85%, still more preferably at least 90%, and most preferably at least 95%. The polypeptide products of homologous genes have at least 35% amino acid sequence identity over at least one sequence window of 18 amino acid residues, more preferably at least 40%, still more preferably at least 50% or 60%, and most preferably at least 70%, 80%, or 90%. Preferably, the homologous gene product is also a functional homolog, meaning that the homolog will functionally complement one or more biological activities of the product being compared. For nucleotide or amino acid sequence comparisons where a homology is defined by a % sequence identity, the percentage is determined using BLAST programs (with default parameters (Altschul et al., 1997, "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acid Res. 25:3389-3402). Any of a variety of algorithms known in the art which provide comparable results can also be used, preferably using default parameters. Performance characteristics for three different algorithms in homology searching is described in Salamov et al., 1999, "Combining sensitive database searches with multiple intermediates to detect distant homologues." *Protein Eng.* 12:95-100. Another exemplary program package is the GCG™ package from the University of Wisconsin.

Homologs may also or in addition be characterized by the ability of two complementary nucleic acid strands to hybridize to each other under appropriately stringent conditions. Hybridizations are typically and preferably conducted with probe-length nucleic acid molecules, preferably 20-100 nucleotides in length. Those skilled in the art understand how to estimate and adjust the stringency of hybridization conditions such that sequences having at least a desired level of complementarity will stably hybridize, while those having lower complementarity will not. For examples of hybridization conditions and parameters, see, *e.g.,.* Maniatis, T. et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor University Press, Cold Spring, N.Y.; Ausubel, F.M. et al. (1994) Current Protocols in Molecular Biology.

SD-63311.4

John Wiley & Sons, Secaucus, N.J. Homologs and homologous gene sequences may thus be identified using any nucleic acid sequence of interest, including the phage ORFs and bacterial target genes of the present invention.

A typical hybridization, for example, utilizes, besides the labeled probe of interest, a salt solution such as 6xSSC (NaCl and Sodium Citrate base) to stabilize nucleic acid strand interaction, a mild detergent such as 0.5% SDS, together with other typical additives such as Denhardt's solution and salmon sperm DNA. The solution is added to the immobilized sequence to be probed and incubated at suitable temperatures to preferably permit specific binding while minimizing nonspecific binding. The temperature of the incubations and ensuing washes is critical to the success and clarity of the hybridization. Stringent conditions employ relatively higher temperatures, lower salt concentrations, and/or more detergent than do non-stringent conditions. Hybridization temperatures also depend on the length, complementarity level, and nature (ie, "GC content") of the sequences to be tested. Typical stringent hybridizations and washes are conducted at temperatures of at least 40°C, while lower stringency hybridizations and washes are typically conducted at 37°C down to room temperature (~25°C). One of skill in the art is aware that these conditions may vary according to the parameters indicated above, and that certain additives such as formamide and dextran sulphate may also be added to affect the conditions.

By "stringent hybridization conditions" is meant hybridization conditions at least as stringent as the following: hybridization in 50% formamide, 5X SSC, 50 mM NaH$_2$PO$_4$, pH 6.8, 0.5% SDS, 0.1 mg/mL sonicated salmon sperm DNA, and 5X Denhart's solution at 42°C overnight; washing with 2X SSC, 0.1% SDS at 45°C; and washing with 0.2X SSC, 0.1% SDS at 45°C.

In sequence comparison analyses, an ORF, or motif, or set of motifs in a bacteriophage sequence can be compared to known inhibitor sequences, e.g., homologous sequences encoding homologous inhibitors of bacterial function. Likewise, the analysis can include comparison with the structure of essential bacterial gene products, as structural similarities can be indicative of similar or replacement biological function. Such analysis can include the identification of a signature, or characteristic motif(s) of an inhibitor or inhibitor class.

Also, the identification of structural motifs in an encoded product, based on nucleotide or amino acid sequence analysis, can be used to infer a biochemical function for the product. A database containing identified structural motifs in a large number of sequences is available for identification of motifs in phage sequences. The database is PROSITE, which is available at www.expasy.ch/cgi~bin/scanprosite. The identification of motifs can, for example, include the identification of signature motifs for a class or classes of inhibitory proteins. Other such databases may also be used.

SD-63311.4

In aspects and preferred embodiments described herein, in which a bacterium or host bacterium is specified, the bacterium or host bacterium is preferably selected from a pathogenic bacterial species, for example, one selected from Table 1. Preferably, an animal or plant pathogen is used. For animals, preferably the bacterium is a bird or mammalian pathogen, still more preferably a human pathogen.

In aspects and preferred embodiments involving a bacteriophage or sequences from a bacteriophage, one or more bacteriophage are preferably selected from those listed in Table 1 in the Detailed Description below. Those exemplary bacteriophge are readily obtained from the indicated sources.

In some cases, it is advantageous to utilize phage with non-pathogenic host bacteria. The genome, structural motif, ORF, homolog, and other analyses described herein can be performed on such phage and bacteria. Such analysis provides useful information and compositions. The results of such analyses can also be utilized in aspects of the present invention to identify homologous ORFs, especially inhibitor ORFs in phage with pathogenic bacterial hosts. Similarly, identification of a target in a non-pathogenic host can be used to identify homologous sequences and targets in pathogenic bacteria, especially in genetically closely related bacteria. Those skilled in the art are familiar with bacterial genetic relationships and with how to determine relatedness based on levels of genomic identity or other measures of nucleotide sequence and/or amino acid sequence similarity, and/or other physical and culture characteristics such as morphology, nutritional requirements, or minimal media to support growth.

Also in preferred embodiments, an embodiments of this aspect is combined with an embodiment of the following aspect.

A related aspect of the invention provides methods for identifying a target for antibacterial agents by identifying the bacterial target(s) of at least one uncharacterized or untargeted inhibitor protein or RNA from a bacteriophage. Such identification allows the development of antibacterial agents active on such targets. Preferred embodiments for identifying such targets involve the identification of binding of target and phage ORF products to one another. The phage ORF products may be subportions of a larger ORF product that also binds the host target. In preferred embodiments, the phage protein or RNA is from an uncharacterized bacteriophage in Table 1. This aspect preferably includes the identification of a plurality of such targets in one or a plurality of different bacteria, preferably in one or a plurality of bacteria listed in Table 1.

In preferred embodiments of this aspect and other aspects of this invention involving particular phage ORFs or phage sequences, the ORF is *Staphylococcus*

aureus phage 77 ORF 17, 19, 43, 102, 104, or 182 as identified in U.S. application 09/407,804.

As indicated for the above aspect, preferably the method involves the use of a plurality of different phage, and thus a plurality of different phage inhibitors and/or

5   inhibitor ORFs.

In addition to uncharacteized phage ORF products, it is also useful to identify the targets of phage ORF products which are known to be inhibitors of host bacteria, but where the target has not been identified. Thus, such inhibitors can likewise be utilized as "untargeted" inhibitor phage ORFs and ORF products, *e.g.*, proteins or

10   RNAs.

In the context of inhibitor proteins or RNAs from a phage, the term "uncharacterized" means that a bacteria-inhibiting function for the protein has not previously been identified. Preferably, but not necessarily, the sequence of the protein or the corresponding coding region or ORF was not described in the art before the

15   filing of the present application for patent (or alternatively prior to the present invention). Thus, this term specifically excludes any bacteria-inhibiting phage protein and its associated bacterial target which has been identified as inhibitory before the present invention or alternatively before the filing of the present application, for example those identified in Tables 12-14 or otherwise identified herein. For example,

20   from *E. coli*, phage T7 genes 0.7 and 2.0 target the host RNA polymerase, phage T4 gp55/gp33 alter the specificity of host RNA polymerase. The T4 *regB* gene product also targets the host translation apparatus. As with the uncharacterized bacteriophage ORFs or bacteriophage above, for such identified proteins, the sequences encoding those proteins are excluded from the uncharacterized inhibitor proteins.

25   The term "fragment" refers to a portion of a larger molecule or assembly. For proteins, the term "fragment" refers to a molecule which includes at least 5 contiguous amino acids from the reference polypeptide or protein, preferably at least 8, 10, 12, 15, 20, 30, 50 or more contiguous amino acids. In connection with oligo- or polynucleotides, the term "fragment" refers to a molecule which includes at least 15

30   contiguous nucleotides from a reference polynucleotide, preferably at least 24, 30, 36, 45, 60, 90, 150, or more contiguous nucleotides.

Preferred embodiments involve identification of binding that include methods for distinguishing bound molecules, for example, affinity chromatography, immunoprecipitation, crosslinking, and/or genetic screen methods that permit

35   protein:protein interactions to be monitored. One of skill in the art is familiar with these techniques and common materials utilized (see, *e.g.*, Coligan, J. et al. (eds.) (1995) <u>Current Protocols in Protein Science</u>, John Wiley & Sons, Secaucus, N.J.).

SD-63311.4

Genetic screening for the identification of protein:protein interactions typically involves the co-introduction of both a chimeric bait nucleic acid sequence (here, the phage ORF to be tested) and a chimeric target nucleic acid sequence that, when co-expressed and having affinity for one another in a host cell, stimulate reporter gene

5 expression to indicate the relationship. A "positive" can thus suggest a potential inhibitory effect in bacteria. This is discussed in further detail in the Detailed Description section below. In this way, new bacterial targets can be identified that are inhibited by specific phage ORF products or derivatives, fragments, mimetics, or other molecules.

10 Other embodiments involve the identification and/or utilization of mutant targets by virtue of their host's relatively unresponsive nature in the presence of expression of ORFs previously identified as inhibitory to the non-mutant or wild-type strain. Such mutants have the effect of protecting the host from an inhibition that would otherwise occur and indirectly allow identification of the precise responsible

15 target for follow-up studies and anti-microbial development. In certain embodiments, rescue from inhibition occurs under conditions in which a bacterial target or mutant target is highly expressed. This is performed, for example, through coupling of the sequence with regulatory element promoters, e.g., as known in the art, which regulate expression at levels higher than wild-type, e.g., at a level sufficiently higher that the

20 inhibitor can be competitively bound to the highly expressed target such that the bacterium is detectably less inhibited.

Identification of the bacterial target can involve identification of a phage-specific site of action. This can involve a newly identified target, or a target where the phage site of action differs from the site of action of a previously known antibacterial

25 agent or inhibitor. For example, phage T7 genes 0.7 and 2.0 target the host RNA polymerase, which is also the cellular target for the antibacterial agent, rifampin. To the extent that a phage product is found to act at a different site than previously described inhibitors, aspects of the present invention can utilize those new, phage-specific sites for identification and use of new agents. The site of action can be

30 identified by techniques well-known to those skilled in the art, for example, by mutational analysis, binding competition analysis, and/or other appropriate techniques.

Once a bacterial host target protein or nucleic acid or mutant target sequence has been identified and/or isolated, it too can be conveniently sequenced, sequence

35 analyzed (e.g., by computer), and the underlying gene(s), and corresponding translated product(s) further characterized. Preferred embodiments include such analysis and identification. Preferably such a target has not previously been identified as an appropriate target for antibacterial action.

SD-63311.4

        Certain embodiments include the identification of at least one inhibitory phage
ORF or ORF product, *e.g.*, as described for the above aspect, and thus are a
combination of the two aspects.

        Additionally, the invention provides methods for identifying targets for
antibacterial agents by identifying homologs of a *Enterococcus* sp. target of a
bacteriophage inhibitory ORF product. Such homologs may be utilized in the various
aspects and embodiments described herein as describded for the host *Enterococcus* sp.
for bacteriophage 182.

        Other aspects of the invention provide isolated, purified, or enriched specific
phage nucleic acid and amino acid sequences, subsequences, and homologs thereof for
phage selected from uncharacterized phage listed in Table 1, preferably from
bacteriophage 77, 3A, 96. For example, such sequences do not include sequences
identified in any of Tables 11-14. Such nucleotide sequences are at least 15
nucleotides in length, preferably at least 18, 21, 24, or 27 nucleotides in length, more
preferably at least 30, 50, or 90 nucleotides in length. In certain embodiments, longer
nucleic acids are preferred, for example those of at least 120, 150, 200, 300, 600, 900
or more nucleotides. Such sequences can, for example, be amplification
oligonucleotides (*e.g.*, PCR primers), oligonucleotide probes, sequences encoding a
portion or all of a phage-encoded protein, or a fragment or all of a phage-encoded
protein. In preferred embodiments, the nucleic acid sequence contains a sequence
which is within a length range with a lower length as specified above, and an upper
length limit which is no more than 50, 60, 70, 80, or 90% of the length of the
corresponding full-length ORF. The upper length limit can also be expressed in terms
of the number of base pairs of the ORF (coding region). In preferred embodiments,
the nucleic acid sequence is from *Staphylococcus aureus* phage 77 ORF 17, 19, 43,
102, 104, or 182 as identified in U.S. application 09/407,804.

        As it is recognized that alternate codons will encode the same amino acid for
most amino acids due to the degeneracy of the genetic code, the sequences of this
aspect includes nucleic acid sequences utilizing such alternate codon usage for one or
more codons of a coding sequence. For example, all four nucleic acid sequences
GCT, GCC, GCA, and GCG encode the amino acid, alanine. Therefore, if for an
amino acid there exists an average of three codons, a polypeptide of 100 amino acids
in length will, on average, be encoded by 3^{100}, or 5×10^{47}, nucleic acid sequences.
Thus, a nucleic acid sequence can be modified (*e.g.*, a nucleic acid sequence from a
phage as specified above) to form a second nucleic acid sequence encoding the same
polypeptide as encoded by the first nucleic acid sequence using routine procedures
and without undue experimentation. Thus, all possible nucleic acid sequences that
encode the specified amino acid sequences are also fully described herein, as if all

SD-63311.4

were written out in full, taking into account the codon usage, especially that preferred in the host bacterium. The alternate codon descriptions are available in common texbooks, for example, Stryer, BIOCHEMISTRY 3rd ed., and Lehninger, BIOCHEMISTRY 3rd ed. Codon preference tables for various types of organisms are

5   available in the literature. Sequences with alternate codons at one or more sites can also be utilized in the computer-related aspects and embodiments herein. Because of the number of sequence variations involving alternate codon usage, for the sake of brevity, individual sequences are not separately listed herein. Instead the alternate sequences are described by reference to the natural sequence with replacement of one

10   or more (up to all) of the degenerate codons with alternate codons from the alternate codon table (Table 6), preferably with selection according to preferred codon usage for the normal host organism or a host organism in which a sequence is intended to be expressed. Those skilled in the art also understand how to alter the alternate codons to be used for expression in organisms where certain codons code differently than shown

15   in the "universal" codon table.

        For amino acid sequences or polypeptides, sequences contain at least 5 peptide-linked amino acid residues, and preferably at least 6, 7, 10, 15, 20, 30, or 40, amino acids having identical amino acid sequence as the same number of contiguous amino acid residues in a particular phage ORF product. In some cases longer sequences may

20   be preferred, for example, those of at least 50, 60, 70, 80, or 100 amino acids in length. In preferred embodiments, the amino acid sequence contains a sequence which is within a length range with a lower length as specified above, and an upper length limit which is no more than 50, 60, 70, 80, or 90% of the length of the corresponding full-length ORF product. The upper length limit can also be expressed in terms of the

25   number of amino acid residues of the ORF product. In preferred embodiments, the amino acid sequence or polypeptide has bacteria-inhibiting function when expressed or otherwise present in a bacterial cell that is a host for the bacteriophage from which the sequence was derived.

        By "isolated" in reference to a nucleic acid is meant that a naturally occurring

30   sequence has been removed from its normal cellular (e.g., chromosomal) environment or is synthesized in a non-natural environment (e.g., artificially synthesized). Thus, the sequence may be in a cell-free solution or placed in a different cellular environment. The term does not imply that the sequence is the only nucleotide chain present, but that it is essentially free (about 90-95% pure at least) of non-nucleotide

35   material naturally associated with it, and thus is distinguished from isolated chromosomes.

        The term "enriched" means that the specific DNA or RNA sequence constitutes a significantly higher fraction (2-5 fold) of the total DNA or RNA present

SD-63311.4

in the cells or solution of interest than in normal or diseased cells or in cells from which the sequence was originally taken. This could be caused by a person by preferential reduction in the amount of other DNA or RNA present, or by a preferential increase in the amount of the specific DNA or RNA sequence, or by a

5   combination of the two. However, it should be noted that enriched does not imply that there are no other DNA or RNA sequences present, just that the relative amount of the sequence of interest has been significantly increased.

The term "significant" is used to indicate that the level of increase is useful to the person making such an increase and an increase relative to other nucleic acids of

10   about at least 2-fold, more preferably at least 5- to 10-fold or even more. The term also does not imply that there is no DNA or RNA from other sources. The other source DNA may, for example, comprise DNA from a yeast or bacterial genome, or a cloning vector such as pUC19. This term distinguishes from naturally occurring events, such as viral infection, or tumor type growths, in which the level of one

15   mRNA may be naturally increased relative to other species of mRNA. That is, the term is meant to cover only those situations in which a person has intervened to elevate the proportion of the desired nucleic acid.

It is also advantageous for some purposes that a nucleotide sequence be in purified form. The term "purified" in reference to nucleic acid does not require

20   absolute purity (such as a homogeneous preparation). Instead, it represents an indication that the sequence is relatively more pure than in the natural environment (compared to the natural level, this level should be at least 2-5 fold greater, *e.g.*, in terms of mg/mL). Individual clones isolated from a cDNA library may be purified to electrophoretic homogeneity. The claimed DNA molecules obtained from these

25   clones could be obtained directly from total DNA or from total RNA. The cDNA clones are not naturally occurring, but rather are preferably obtained via manipulation of a partially purified naturally occurring substance (messenger RNA). The construction of a cDNA library from mRNA involves the creation of a synthetic substance (cDNA) and pure individual cDNA clones can be isolated from the

30   synthetic library by clonal selection of the cells carrying the cDNA library. Thus, the process which includes the construction of a cDNA library from mRNA and isolation of distinct cDNA clones yields an approximately 10^6-fold purification of the native message. Thus, purification of at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly

35   contemplated.

The terms "isolated", "enriched", and "purified" as used with respect to nucleic acids, above, may similarly be used to denote the relative purity and abundance of polypeptides (multimers of amino acids joined one to another by α-

SD-63311.4

carboxyl:α-amino group (peptide) bonds). These, too, may be stored in, grown in, screened in, and selected from libraries using biochemical techniques familiar in the art. Such polypeptides may be natural, synthetic or chimeric and may be extracted using any of a variety of methods, such as antibody immunoprecipitation, other

5   "tagging" techniques, conventional chromatography and/or electrophoretic methods. Some of the above utilize the corresponding nucleic acid sequence.

As indicated above, aspects and embodiments of the invention are not limited to entire genes and proteins. The invention also provides and utilizes fragments and portions thereof, preferably those which are "active" in the inhibitory sense described

10   above. Such peptides or oligopeptides and oligo or polynucleotides have preferred lengths as specified above for nucleic acid and amino acid sequences from phage; corresponding recombinant constructs can be made to express the encoded same. Also included are homologous sequences and fragments thereof.

The nucleotide and amino acid sequences identified herein are believed to be

15   correct, however, certain sequences may contain a small percentage of errors, e.g., 1-5%. In the event that any of the sequences have errors, the corrected sequences can be readily provided by one skilled in the art using routine methods. For example, the nucleotide sequences can be confirmed or corrected by obtaining and culturing the relevant phage, and purifying phage genomic nucleic acids. A region or regions of

20   interest can be amplified, e.g., by PCR from the appropriate genomic template, using primers based on the described sequence. The amplified regions can then be sequenced using any of the available methods (e.g., a dideoxy termination method). This can be done redundantly to provide the corrected sequence or to confirm that the described sequence is correct. Alternatively, a particular sequence or sequences can

25   be identified and isolated as an insert or inserts in a phage genomic library and isolated, amplified, and sequenced by standard methods. Confirmation or correction of a nucleotide sequence for a phage gene provides an amino acid sequence of the encoded product by merely reading off the amino acid sequence according to the normal codon relationships and/or expressed in a standard expression system and the

30   polypeptide product sequenced by standard techniques. The sequences described herein thus provide unique identification of the corresponding genes and other sequences, allowing those sequences to be used in the various aspects of the present invention.

In other aspects the invention provides recombinant vectors and cells

35   harboring at least one of the phage ORFs or portion thereof, or bacterial target sequences described herein. As understood by those skilled in the art, vectors may be provided in different forms, including, for example, plasmids, cosmids, and virus-based vectors. See, e.g., Maniatis, T. et al. (1989) <u>Molecular Cloning: A Laboratory</u>

SD-63311.4

Manual, Cold Spring Harbor University Press, Cold Spring, N.Y.; See also, Ausubel,
F.M. et al. (eds.) (1994) Current Protocols in Molecular Biology. John Wiley & Sons,
Secaucus, N.J.

5   In preferred embodiments, the vectors will be expression vectors, preferably
shuttle vectors that permit cloning, replication, and expression within bacteria. An
"expression vector" is one having regulatory nucleotide sequences containing
transcriptional and translational regulatory information that controls expression of the
nucleotide sequence in a host cell. Preferably the vector is constructed to allow
amplification from vector sequences flanking an insert locus. In certain embodiments,

10   the expression vectors may additionally or alternativley support expression, and/or
replication in animal, plant and/or yeast cells due to the presence of suitable
regulatory sequences, e.g., promoters, enhancers, 3' stabilizing sequences, primer
sequences, etc. In preferred embodiments, the promoters are inducible and specific
for the system in which expression is desired, e.g., bacteria, animal, plant, or yeast.

15   The vectors may optionally encode a "tag" sequence or sequences to facilitate protein
purification. Convenient restriction enzyme cloning sites and suitable selective
marker(s) are also optionally included. Such selective markers can be, for example,
antibiotic resistance markers or markers which supply an essential nutritive growth
factor to an otherwise deficient mutant host, e.g., tryptophan, histidine, or leucine in

20   the Yeast Two-Hybrid systems described below.

The term "recombinant vector" relates to a single- or double-stranded circular
nucleic acid molecule that can be transfected into cells and replicated within or
independently of a cell genome. A circular double-stranded nucleic acid molecule can
be cut and thereby linearized upon treatment with appropriate restriction enzymes. An

25   assortment of nucleic acid vectors, restriction enzymes, and the knowledge of the
nucleotide sequences cut by restriction enzymes are readily available to those skilled
in the art. A nucleic acid molecule encoding a desired product can be inserted into a
vector by cutting the vector with restriction enzymes and ligating the two pieces
together. Preferably the vector is an expression vector, e.g., a shuttle expression

30   vector as described above.

By " recombinant cell" is meant a cell possessing introduced or engineered
nucleic acid sequences, e.g., as described above. The sequence may be in the form of
or part of a vector or may be integrated into the host cell genome. Preferably the cell
is a bacterial cell.

35   In another aspect, the invention also provides methods for identifying and/or
screening compounds "active on" at least one bacterial target of a bacteriophage
inhibitor protein or RNA. Preferred embodiments involve contacting such a bacterial
target or targets (e.g., bacterial target proteins) with a test compound, and determining

SD-63311.4

whether the compound binds to or reduces the level of activity of the bacterial target (*e.g.*, a bacterial target protein). Preferably this is done either *in vivo* (*i.e.*, in a cell-based assay) or *in vitro*, e.g., in a cell-free system under approximately physiological conditions.

5   The compounds that can be used may be large or small, synthetic or natural, organic or inorganic, proteinaceous or non-proteinaceous. In preferred embodiments, the compound is a peptidomimetic, as described herein, a bacteriophage inhibitor protein or fragment or derivative thereof, preferably an "active portion", or a small molecule.

10   In particular embodiments, the methods include the identification of bacterial targets or the site of action of an inhibitor on a bacterial target as described above or otherwise described herein.

In embodiments involving binding assays, preferably binding is to a fragment or portion of a bacterial target protein, where the fragment includes less than 90%,

15   80%, 70%, 60%, 50%, 40%, or 30% of an intact bacterial target protein. Preferably, the at least one bacterial target includes a plurality of different targets of bacteriophage inhibitor proteins, preferably a plurality of different targets. The plurality of targets can be in or from a plurality of different bacteria, but preferably is from a single bacterial species.

20   A "method of screening" refers to a method for evaluating a relevant activity or property of a large plurality of compounds (e.g., a bacteria-inhibiting activity), rather than just one or a few compounds. For example, a method of screening can be used to conveniently test at least 100, more preferably at least 1000, still more preferably at least 10,000, and most preferably at least 100,000 different compounds,

25   or even more.

In the context of this invention, the term "small molecule" refers to compounds having molecular mass of less than 2000 Daltons, preferably less than 1500, still more preferably less than 1000, and most preferably less than 600 Daltons. Preferably but not necessarily, a small molecule is not an oligopeptide.

30   In a related aspect or in preferred embodiments, the invention provides a method of screening for potential antibacterial agents by determining whether any of a plurality of compounds, preferably a plurality of small molecules, is active on at least one target of a bacteriophage inhibitor protein or RNA. Preferred embodiments include those described for the above aspect, including embodiments which involve

35   determining whether one or more test compounds bind to or reduce the level of activity of a bacterial target, and embodiments which utilize a plurality of different targets as described above.

The identification of bacteria-inhibiting phage ORFs and their encoded products also provides a method for identifying an active portion of such an encoded product. This also provides a method for identifying a potential antibacterial agent by identifying such an active portion of a phage ORF or ORF product. In preferred

5   embodiments, the identification of an active portion involves one or more of mutational analysis, deletion analysis, or analysis of fragments of such products. The method can also include determination of a 3-dimensional structure of an active portion, such as by analysis of crystal diffraction patterns. In further embodiments, the method involves constructing or synthesizing a peptidomimetic compound, where

10   the structure of the peptidomimetic compound corresponds to the structure of the active portion. In this context, "corresponds" means that the peptidomimetic compound structure has sufficient similarities to the structure of the active portion that the peptidomimetic will interact with the same molecule as the phage protein and preferably will elicit at least one cellular response in common which relates to the

15   inhibition of the cell by the phage protein.

The methods for identifying or screening for compounds or agents active on a bacterial target of a phage-encoded inhibitor can also involve identification of a phage-specific site of action on the target.

Preferably in the methods for identifying or screening for compounds active

20   on such a bacterial target, the target is uncharacterized; the target is from an uncharacterized bacterium from Table 1; the site of action is a phage-specfic site of action.

Further embodiments include the identification of inhibitor phage ORFs and bacterial targets as in aspects above.

25   An "active portion" as used herein denotes an epitope, a catalytic or regulatory domain, or a fragment of a bacteriophage inhibitor protein that is responsible for, or a significant factor in, bacterial target inhibition. The active portion preferably may be removed from its contiguous sequences and, in isolation, still effect inhibition.

By "mimetic" is meant a compound structurally and functionally related to a

30   reference compound that can be natural, synthetic, or chimeric. In terms of the present invention, a "peptidomimetic," for example, is a compound that mimics the activity-related aspects of the 3-dimensional structure of a peptide or polyeptide in a non-peptide compound, for example mimics the structure of a peptide or active portion of a phage- or bacterial ORF-encoded polypeptide.

35   A related aspect provides a method for inhibiting a bacterial cell by contacting the bacterial cell with a compound active on a bacterial target of a bacteriophage inhibitor protein or RNA, where the target was uncharacterized. In preferred embodiments, the compound is such a protein, or a fragment or derivative thereof; a

SD-63311.4

structural mimetic, *e.g.*, a peptidomimetic, of such a protein or fragment; a small molecule; the contacting is performed *in vitro*, the contacting is performed *in vivo* in an infected or at risk organism, *e.g.*, an animal such as a mammal or bird, for example, a human, or other mammal described herein; the bacterium is selected from a genus

5   and/or species listed in Table 1; the bacteriophage inhibitor protein is uncharacterized; and the bacteriophage inhibitor protein is from an uncharacterized phage listed in Table 1.

In the context of targets in this invention, the term "uncharacterized" means that the target was not recognized as an appropriate target for an antibacterial agent

10 prior to the filing of the present application or alternatively prior to the present invention. Such lack of recognition can include, for example, situations where the target and/or a nucleotide sequence encoding the target were unknown, situations where the target was known, but where it had not been identified as an appropriate target or as an essential cellular component, and situations where the target was

15 known as essential but had not been recognized as an appropriate target due to a belief that the target would be inaccessible or otherwise that contacting the cell with a compound active on the target *in vitro* would be ineffective in cellular inhibition, or ineffective in treatment of an infection. Methods described herein utilizing bacterial targets, *e.g.*, for inhibiting bacteria or treating bacterial infections, can also utilize

20 "uncharacterized target sites", meaning that the target has been previously recognized as an appropriate target for an antibacterial agent, but where an agent or inhibitor of the invention is used which acts at a different site than that at which the previously utilized antibacterial agent, *i.e.*, a phage-specific site. Preferably the phage-specific site has different functional characteristics from the previously utilized site. In the

25 context of targets or target sites, the term "phage-specific" indicates that the target or site is utilized by at least one bacteriophage as an inhibitory target and is different from previously identified targets or target sites.

In the context of this invention, the term "bacteriophage inhibitor protein" refers to a protein encoded by a bacteriophage nucleic acid sequence which inhibits

30 bacterial function in a host bacterium. Thus, it is a bacteria-inhibiting phage product.

In the context of this invention, the phrase "contacting the bacterial cell with a compound active on a bacterial target of a bacteriophage inhibitor protein" or equivalent phrases refer to contacting with an isolated, purified, or enriched compound or a composition including such a compound, but specifically does not rely

35 on contacting the bacterial cell with an intact phage which encodes the compound. Preferably no intact phage are involved in the contacting.

Related aspects provide methods for prophylactic or therapeutic treatment of a bacterial infection by administering to an infected, challenged or at risk organism a

SD-63311.4

therapeutically or prophylactically effective amount of a compound active on a target of a bacteriophage inhibitor protein or RNA, or as described for the previous aspect. Preferably the bacterium involved in the infection or risk of infection produces the identified target of the bacteriophage inhibitor protein or alternatively produces a homologous target compound. In preferred embodiments, the host organism is a plant or animal, preferably a mammal or bird, and more preferably, a human or other mammal described herein. Preferred embodiments include, without limitation, those as described for the preceding aspect.

Compounds useful for the methods of inhibiting, methods of treating, and pharmaceutical compositions can include novel compounds, but can also include compounds which had previously been identified for a purpose other than inhibition of bacteria. Such compounds can be utilized as described and can be included in pharmaceutical compositions.

In preferred embodiments of this and other aspects of the invention utilizing bacterial target sequences of a bacteriiophage inhibitory ORF product, the target sequence is encoded by a *Staphylococcus* nucleic acid coding sequence, preferably *S. aureus*. Possible target sequences are described herein by reference to sequence source sites.

The amino acid sequence of a polypeptide target is readily provided by translating the corresponding coding region. For the sake of brevity, the sequences are not reproduced herein. For the sake of brevity, the sequences are described by reference to the GenBank entries instead of being written out in full herein. In cases where the TIGR or GenBank entry for a coding region is not complete, the complete sequence can be readily obtained by routine methods, e.g., by isolating a clone in a phage host genomic library, and sequencing the clone insert to provide the relevant coding region. The boundaries of the coding region can be identified by conventional sequence analysis and/or by expression in a bacterium in which the endogenous copy of the coding region has been inactivated and using subcloning to identify the functional start and stop codons for the coding region.

In the context of nucleic acid or amino acid sequences of this invention, the term "corresponding" indicates that the sequence is at least 95% identical, preferably at least 97% identical, and more preferably at least 99% identical to a sequence from the specified phage genome, a ribonucleotide equivalent, a degenerate equivalent (utilizing one or more degenerate codons), or a homologous sequence, where the homolog provides functionally equivalent biological function.

By "treatment" or "treating" is meant administering a compound or pharmaceutical composition for prophylactic and/or therapeutic purposes. The term "prophylactic treatment" refers to treating a patient or animal that is not yet infected

SD-63311.4

but is susceptible to or otherwise at risk of a bacterial infection. The term "therapeutic treatment" refers to administering treatment to a patient already suffering from infection.

The term "bacterial infection" refers to the invasion of the host organism, animal or plant, by pathogenic bacteria. This includes the excessive growth of bacteria which are normally present in or on the body of the organism, but more generally, a bacterial infection can be any situation in which the presence of a bacterial population(s) is damaging to a host organism. Thus, for example, an organism suffers from a bacterial population when excessive numbers of a bacterial population are present in or on the organism's body, or when the effects of the presence of a bacterial population(s) is damaging to the cells, tissue, or organs of the organism.

The terms "administer", "administering", and "administration" refer to a method of giving a dosage of a compound or composition, *e.g.*, an antibacterial pharmaceutical composition, to an organism. Where the organism is a mammal, the method is, *e.g.*, topical, oral, intravenous, transdermal, intraperitoneal, intramuscular, or intrathecal. The preferred method of administration can vary depending on various factors, *e.g.*, the components of the pharmaceutical composition, the site of the potential or actual bacterial infection, the bacterium involved, and the infection severity.

The term "mammal" has its usual biological meaning referring to any organism of the Class Mammalia of higher vertebrates that nourish their young with milk secreted by mammary glands, *e.g.*, mouse, rat, and, in particular, human, bovine, sheep, swine, dog, and cat.

In the context of treating a bacterial infection a "therapeutically effective amount" or "pharmaceutically effective amount" indicates an amount of an antibacterial agent, *e.g.*, as disclosed for this invention, which has a therapeutic effect. This generally refers to the inhibition, to some extent, of the normal cellular functioning of bacterial cells that renders or contributes to bacterial infection.

The dose of antibacterial agent that is useful as a treatment is a "therapeutically effective amount." Thus, as used herein, a therapeutically effective amount means an amount of an antibacterial agent that produces the desired therapeutic effect as judged by clinical trial results and/or animal models. This amount can be routinely determined by one skilled in the art and will vary depending on several factors, such as the particular bacterial strain involved and the particular antibacterial agent used.

In connection with claims to methods of inhibiting bacteria and therapeutic or prophylactic treatments, "a compound active on a target of a bacteriophage inhibitor protein" or terms of equivalent meaning differ from administration of or contact with

SD-63311.4

an intact phage naturally encoding the full-length inhibitor compound. While an intact phage may conceivably be incorporated in the present methods, the method at least includes the use of an active compound as specified different from a full length inhibitor protein naturally encoded by a bacteriophage and/or a delivery or contacting

5   method different from administration of or contact with an intact phage encoding the full-length protein. Similarly, pharmaceutical compositions described herein at least include an active compound different from a full-length inhibitor protein naturally encoded by a bacteriophage or such a full-length protein is provided in the composition in a form different from being encoded by an intact phage. Preferably

10   the methods and compositions do not include an intact phage.

In accord with the above aspects, the invention also provides antibacterial agents and compounds active on bacterial targets of bacteriophage inhibitor proteins or RNAs, where the target was uncharacterized as indicated above. As previously indicated, such active compounds include both novel compounds and compounds

15   which had previously been identified for a purpose other than inhibition of bacteria. Such previously identified biologically active compounds can be used in embodiments of the above methods of inhibiting and treating. In preferred embodiments, the targets, bacteriophage, and active compound are as described herein for methods of inhibiting and methods of treating. Preferably the agent or compound

20   is formulated in a pharmaceutical composition which includes a pharmaceutically acceptable carrier, excipient, or diluent. In addition, the invention provides agents, compounds, and pharmaceutical compositions where an active compound is active on an uncharacterized phage-specific site.

In preferred embodiments, the target is as described for embodiments of

25   aspects above.

Likewise, the invention provides a method of making an antibacterial agent. The method involves identifying a target of a bacteriophage inhibitor polypeptide or protein or RNA, screening a plurality of compounds to identify a compound active on the target, and synthesizing the compound in an amount sufficient to provide a

30   therapeutic effect when administered to an organism infected by a bacterium naturally producing the target. In preferred embodiments, the identification of the target and identification of active compounds include steps or methods and/or components as described above (or otherwise herein) for such identification. Likewise, the active compound can be as described above, including fragments and derivatives of phage

35   inhibitor proteins, peptidomimetics, and small molecules. As recognized by those skilled in the art, peptides can be synthesized by expression systems and purified, or can be synthesized artificially.

As indicated above, sequence analysis of nucleotide and/or amino acid sequences can beneficially utilize computer analysis. Thus, in additional aspects the invention provides computer-related hardware and media and methods utilizing and incorporating sequence data from uncharacterized phage, *e.g.*, uncharacterized phage

5   listed in Table 1, preferably at least one of bacteriophage 77, 3A, and 96, (*Staphylococcus aureus* phage). In general, such aspects can facilitate the above described aspects. Various embodiments involve the analysis of genetic sequence and encoded products, as applied to the evaluating bacteriophage inhibitor ORFs and compounds and fragments related thereto. The various sequence analyses, as well as

10   function analyses, can be used separately or in combination, as well as in preceding aspects and embodiments. Use in combination is often advantageous as the additional information allows more efficient prioritizing of phage ORFs for identification of those ORFs that provide bacteria-inhibiting function.

In one aspect, the invention provides a computer-readable device which

15   includes at least one recorded amino acid or nucleotide sequence corresponding to one of the specified phage and a sequence analysis program for analyzing a nucleotide and/or amino acid sequence. The device is arranged such that the sequence information can be retrieved and analyzed using the analysis program. The analysis can identify, for example, homologous sequences or the indicated %s of the phage

20   genome and structural motifs. Preferably the sequence includes at least 1 phage ORF or encoded product, more preferably at least 10%, 20%, 30%, 40%, 50%, 70%, 90%, or 100% of the genomic phage ORFs and/or equivalent cDNA, RNA, or amino acid sequences. Preferably the sequence or sequences in the device are recorded in a medium such as a floppy disk, a computer hard drive, an optical disk, computer

25   random access memory (RAM), or magnetic tape. The program may also be recorded in such medium. The sequences can also include sequences from a plurality of different phage.

In this context, the term "corresponding" indicates that the sequence is at least 95% identical, preferably at least 97% identical, and more preferably at least 99%

30   identical to a sequence from the specified phage genome, a ribonucleotide equivalent, a degenerate equivalent (utilizing one or more degenerate codons), or a homologous sequence, where the homolog provides functionally equivalent biological function.

Similarly, the invention provides a computer analysis system for identifying biologically important portions of a bacteriophage genome. The system includes a

35   data storage medium, *e.g.*, as identified above, which has recorded thereon a nucleotide sequence corresponding to at least a portion of at least one uncharacterized bacteriophage genome, a set of program instructions to allow searching of the sequence or sequences to analyze the sequence, and an output device where the

SD-63311.4

portion includes at least the sequence length as specified in the preceding aspect. The output device is preferably a printer, a video display, or a recording medium. More one than one output device may be included. For each of the present computer-related asepcts, the bacteriophage are preferably selected from the uncharacterized phage

5   listed in Table 1, more preferably from bacteriophage 77, 3A, and 96.

In keeping with the computer device aspects, the invention also provides a method for identifying or characterizing a bacteriophage ORF by providing a computer-based system for analyzing nucleotide or amino acid sequences, *e.g.*, as describe above. The system includes a data storage medium which has recorded a

10   sequences or sequences as described for the above devices, a set of instructions as in the preceding aspect, and an output device as in the preceding aspect. The method further involves analyzing at least one sequence, and outputting the analysis results to at least one output device.

In preferred embodiments, the analysis identifies a sequence similarity or

15   homology with a sequence or sequences selected from bacterial ORFs encoding products with related biological function; ORFs encoding known inhibitors; and essential bacterial ORFs. Preferably the analysis identifies a probable biological function based on identification of structural elements or characteristic or signature motifs of an encoded product or on sequence similarity or homology. Preferably the

20   uncharacterized bacteriophage is from Table 1, more preferably at least one of bacteriophage 77, 3A, and 96. In preferred embodiments, the method also involves determining at least a portion of the nucleotide sequence of at least one uncharacterized bacteriophage as indicated, and recording that sequence on data storage medium of the computer-based system.

25

As used in the claims to describe the various inventive aspects and embodiments, "comprising" means including, but not limited to, whatever follows the word "comprising". Thus, use of the term "comprising" indicates that the listed elements are required or mandatory, but that other elements are optional and may or

30   may not be present. By "consisting of" is meant including, and limited to, whatever follows the phrase "consisting of". Thus, the phrase "consisting of" indicates that the listed elements are required or mandatory, and that no other elements may be present. By "consisting essentially of" is meant including any elements listed after the phrase, and limited to other elements that do not interfere with or contribute to the activity or

35   action specified in the disclosure for the listed elements. Thus, the phrase "consisting essentially of" indicates that the listed elements are required or mandatory, but that other elements are optional and may or may not be present depending upon whether or not they affect the activity or action of the listed elements.

SD-63311.4

Further embodiments will be apparent from the following Detailed Description and from the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

5

FIGURE 1A and 1B are flow schematics showing the manipulations necessary to convert pT0021, an arsenite inducible vector containing the luciferase gene, into pTHA or pTM, two *ars* inducible vectors. Vector pTHA contains BamH I, Sal I, and Hind III cloning sites and a downstream HA epitope tag. Vector pTM contains Bam

10   HI and Hind III cloning sites and no HA epitope tag.

FIGURE 2 is a schematic representation of the cloning steps involved to place the DNA segments of any of ORFs 17/ 19/ 43/ 102/104/182 or other sequences into pTHA to assess inhibitory potential. For subcloning into pTM or pT0021, Individual

15   ORFs were amplified by the PCR using oligonucleotides targeting the ATG and stop codons of the ORFs. Using this strategy, Bam HI and Hind III sites were positioned immediately upstream or downstream, respectively of the start and stop codons of each ORF. Following digestion with Bam HI and Hind III, the PCR fragments were subcloned into the same sites of pT0021 or pTM. Clones were verified by PCR and

20   direct sequencing.

FIGURE 3 shows a schematic representation of the functional assays used to characterize the bactericidal and bacteriostatic potential of all predicted ORFs (>33 amino acids) encoded by bacteriophage 77. Fig. 3A) Functional assay on semi-solid

25   support media. Fig. 3B) Functional assay in liquid culture.

FIGURE 4A, B, and C is a bar graph showing the results of a screen in liquid media to assess bacteriostatic or bactericidal activity of 93 predicted ORFs (>33 amino acids) encoded by bacteriophage 77. Growth inhibition assays were performed as

30   detailed in the Detailed Description. The relative growth of Staphylococcus aureus transformants harboring a given bacteriophage 77 ORF (identified on the bottom of the graph), in the absence or presence of arsenite, is plotted relative to growth of a Staphylococcus aureus transformant containing ORF 5, a non-toxic bacteriophage 77

ORF (which is set at 100%). Each bar represents the average obtained from three Staph A transformants grown in duplicate. Bacteriophage 77 ORFs showing significant growth inhibition are plotted in red and consist or ORF 17, 19, 102, 104, and 182.

5

FIGURE 5 shows a block diagram of major components of a general purpose computer.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS
10

The invention may be more clearly understood from the following description. The tables will first be briefly described.

Table 1 is a listing of a large number of available bacteriophage that can be readily obtained and used in the present invention.

15   Table 2 shows the complete nucleotide sequence of the genome of *Staphylococcus aureus* bacteriophage 77.

Table 3 shows a list of all the ORFs from Bacteriophage 77 that were screened in the functional assay to identify those with anti-microbial activity.

Table 4 shows the predicted nucleotide sequence, predicted amino acid
20   sequence, and physiochemical parameters of ORF 17/ 19/ 43/ 102/ 104/ 182]. These include the primary amino acid sequence of the predicted protein, the average molecular weight, amino acid composition, theoretical pI, hydrophobicity map, and predicted secondary structure map.

Table 5 shows homology search results. BLAST analysis was performed with
25   ORFs 17/ 19/ 43/ 102/ 104/ 182 against NCBI non-redundant nucleotide and Swissprot databases. The results of this search indicate that: I) ORF 17 has no significant homology to any gene in the NCBI non-NCBI non-redundant nucleotide database, II) ORF 19 has significant homology to one gene in the NCBI non-redundant nucleotide database - the gene encoding ORF 59 of bacteriophage phi PVL,
30   III) ORF 43 has significant homology to one gene in the NCBI non-redundant nucleotide database - the gene encoding ORF 39 of phi PVL, IV) ORF 102 has significant homology to one gene in the NCBI non-redundant nucleotide database - the gene encoding ORF 38 of phi PVL, V) ORF 104 has no significant homology to

any gene in the NCBI non-redundant nucleotide database, VI) ORF 182 has

significant homology to one gene in the NCBI non-redundant nucleotide database -

the gene encoding ORF 39 of phi PVL.

Table 6 is a table from Alberts et al., MOLECULAR BIOLOGY OF THE

CELL 3rd ed., showing the redundancy of the "universal" genetic code.

Table 7 shows the complete nucleotide sequence of *Staphylococcus aureus*

bacteriophage 3A.

Table 8 is a listing of the ORFs identified in *Staphylococcus aureus*

bacteriophage 3A.

Table 9 shows the complete nucleotide sequence of *Staphylococcus aureus*

bacteriophage 96.

Table 10 is a listing of the ORFs identified in *Staphylococcus aureus*

bacteriophage 96.

Table 11 is a listing of sequences deposited in the NCBI public database

(GeneBank) for bacteriophage listed in Table 1.

Table 12 is a listing of phage which encode a known lysis function , including

the identified lysis gene.

Table 13 is a listing of bacteriophage which encode holin genes, where holin

genes encode proteins which form pores and eventually enable other enzymes to kill

the host bacterium.

Table 14 is a listing of bacteriophage which encode kil genes.

Table 15 is a list of *Staphylococcus aureus* sequences which may include

sequences from genes coding for target sequences for the phage 77-encoded

antimicrobial proteins or peptides.

Background:

As indicated in the Summary above, the present invention is concerned with
the use of bacteriophage coding sequences and the encoded polypeptides or RNA
transcripts to identify bacterial targets for potential new antibacterial agents. Thus,
the invention concerns the selection of relevant bacteria. Particularly relevant bacteria
are those which are pathogens of a complex organism such as an animal, *e.g.*,
mammals, reptiles, and birds, and plants. However, the invention can be applied to
any bacterium (whether pathogenic or not) for which bacteriophage are available or
which are found to have cellular components closely homologous to components
targeted by phage of another bacterium, *e.g.*, a pathogenic bacterium, *e.g.*, a
pathogenic bacterium.

Thus, the invention also concerns the bacteriophage which can infect a selected bacterium. Identification of ORFs or products from the phage which inhibit the host bacterium both provides an inhibitor compound and allows identification of the bacterial target affected by the phage-encoded inhibitor. Such targets are thus

5   identified as potential targets for development of other antibacterial agents or inhibitors and the use of those targets to inhibit those bacteria. As indicated above, even if such a target is not initially identified in a particular bacterium, such a target can still be identified if a homologous target is identified in another bacterium. Usually, but not necessarily, such another bacterium would be a genetically closely

10 related bacterium. Indeed, in some cases, a phage-encoded inhibitor can also inhibit such a homologous bacterial cellular component.

The demonstration that bacteriophage have adapted to inhibiting a host bacterium by acting on a particular cellular component or target provides a strong indication that that component is an appropriate target for developing and using

15 antibacterial agents, e.g., in therapeutic treatments. Thus, the present invention provides additional guidance over mere identification of bacterial essential genes, as the present invention also provides an indication of accessability of the target to an inhibitor, and an indication that the target is sufficiently stable over time (e.g., not subject to high rates of mutation) as phage acting on that target were able to develop

20 and persist. Thus, the present invention identifies a subset of essential cellular components which are particularly likely to be appropriate targets for development of antibacterial agents.

The invention also, therefore, concerns the development or identification of inhibitors of bacteria, in addition to the phage-encoded inhibitory proteins (or RNA

25 transcripts), which are active on the targets of bacteriophage-encoded inhibitors. As described herein, such inhibitors can be of a variety of different types, but are preferably small molecules.

The following description provides preferred methods for developing the various aspects of the invention. However, as those skilled in the art will readily

30 recognize, other approaches can be used to obtain and process relevant information. Thus the invention is not limited to the specifically described methods. In addition, the following description provides a set of steps in a particular order. That series of steps describes the overall development involved in the present invention. However, it is clear that individual steps or portions of steps may be usefully practiced

35 separately, and, further, that certain steps may be performed in a different order or even bypassed if appropriate information is already available or is provided by other sources or methods.

Selecting and Growing Phage, and Isolating DNA

Conceptually, the first step involves selecting bacterial hosts of interest. Preferably, but not necessarily, such hosts will be pathogens of clinical importance. Alternatively, because bacteria all share certain fundamental metabolic and structural features, these features can be targeted for study in one strain, for example a nonpathogenic one, and extrapolated to similarly succeed in pathogenic ones. Nonpathogenic strains may also exhibit initial advantages in being not only less dangerous, but also, for example, in having better growth and culturing characteristics and/or better developed molecular biology techniques and reagents. Consequently, advantageously the invention provides the ability target virtually any bacteria, but preferably pathogenic bacteria, with antimicrobial compounds designed and/or developed using bacteriophage inhibitory proteins and peptides from phage with non-pathogenic and/or pathogenic hosts.

We have selected *Staphylococcus aureus*, *Streptococcus pneumoniae*, various *Enterococci*, and *Pseudomonas aeruginosa* as initial exemplary pathogens. These bacteria are a major cause of morbidity and mortality in hospital-based infections, and the appearance of antibiotics resistance in all three organisms makes it increasingly difficult to treat benign infections involving these organisms. Such infections can include, for example, otitis media, sinusitis, and skin, and airway infections (Neu, H.C. (1992). *Science* 257, 1064-1073). However, the approach described below is clearly applicable to any human bacterial pathogens including but not restricted to *Mycobacterium tuberculosis*, *Nesseria gonorrhoeae*, *Haemophilus influenza*, *Acinobacter*, *Escherichia coli*, *Shigella dysenteria*, *Streptococcus pyogenes*, *Helicobacter pylori*, and *Mycoplasma* species. This invention can also be applied to the discovery of anti-bacterial compounds directed against pathogens of animals other than humans, for example, sheep, cattle, swine, dogs, cats, birds, and reptiles. Similarly, the invention is not limited to animals, but also applies to plants.

The bacteria are grown according to standard methodologies employed in the art, including solid, semi-solid or liquid culturing, which procedures can be found in or extrapolated from standard sources such as Maloy, S.R., Stewart, V.J., and Taylor, R.K. Genetic Analysis of Pathogenic Bacteria (1996) Cold Spring Harbor Laboratory Press, or Maniatis, T. et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor University Press, Cold Spring, N.Y.; or Ausubel, F.M. et al. (1994) Current Protocols in Molecular Biology. John Wiley & Sons, Secaucus, N.J. Culture conditions are selected which are adapted to the particular bacterium generally using culture conditions known in the art as appropriate, or adaptations of those conditions.

Nucleic acids within these bacteria can be routinely extracted through common procedures such as described in the above-referenced manuals and as generally known

SD-63311.4

to those skilled in the art. Those nucleic acid stocks can then be used to practice the other inventive aspects described below.

<u>Selection and Growth of Bacteriophage, and Isolation of DNA</u>

5   The second step involves assembling a group of bacteriophages (phage collection) for each of the targeted bacterial hosts. While the invention can be utilized with a single bacteriophage for a pathogen or other bacterium, it is preferable to utilize a plurality of phage for each bacterium, as comparisons between a plurality of such phage provides useful additional information. Non-limiting examples of phage and

10   sources for some of the above-mentioned pathogenic bacteria are found in Table 1. The criteria used to select such phages is that they are infectious for the microbe targeted, and replicate in, lyse, or otherwise inhibit growth of the bacterium in a measurable fashion. These phages can be very different from one another (representing different families), as judged by criteria such as morphology (head, tail,

15   plate, etc.), and similarity of genome nucleotide sequence (cross-hybridization). Since such diverse bacteriophages are expected to block bacterial host metabolism and ultimately inhibit by a variety of mechanisms, their combined study will lead to the identification of different mechanisms by which the phages independently inhibit bacterial targets. Examples include degradation of host DNA (Parson K.A., and

20   Snustad, D.P. (1975). *J. Virol.* 15, 221-444) and inhibition of host RNA transcription (Severinova, E., Severinov, K. and Darst, S.A. (1998*). J.Mol. Biol.* 279, 9-18). This, in turn, yields novel information on phage proteins that can inhibit the targeted microbe. As explained below, this 1) forms the basis of novel drug discovery efforts based on knowledge of the primary amino acid sequence of the phage inhibitor

25   protein (*e.g.*, peptide fragments or peptidomimetics) and/or 2) leads to the identification of bacterial biochemical pathways, the proteins of which are essential or significant for survival of the targeted microbe, and which enzymatic steps or chemical reactions can be targeted by classical drug discovery methods using molecular inhibitors, for example, small molecule inhibitors.

30   Bacteriophage are generally either of two types, lytic or filamentous, meaning they either outright destroy their host and seek out new hosts after replication, or else continuously propogate and extrude progeny phage from the same host without destroying it. Regardless of the phage life cycle and type, preferred embodiments incorporate phage which impede cell growth in measurable fashion and preferably

35   stop cell growth. To this end, lytic phage are preferred, although certain nonlytic species may also suffice, *e.g.*, if sufficiently bacteriostatic.

Various procedures that are commonly understood by those of skill in the art can be routinely employed to grow, isolate, and purify phage. Such procedures are

SD-63311.4

exemplified by those found in such common laboratory aids such as Maloy, S.R., Stewart, V.J., and Taylor, R.K. Genetic Analysis of Pathogenic Bacteria (1996) Cold Spring Harbor Laboratory Press; Maniatis, T. et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor University Press, Cold Spring, N.Y.; and

5   Ausubel, F.M. et al. (eds.) (1994) Current Protocols in Molecular Biology. John Wiley & Sons, Secaucus, N.J. The techniques generally involve the culturing of infected bacterial cells that are lysed naturally and/or chemically assisted, for example, by the use of an organic solvent such as chloroform that destroys the host cells thereby liberating the phage within. Following this, the cellular debris is

10   centrifuged away from the supernatant containing the phage particles, and the phage then subsequently and selectively precipitated out of the supernatant using various methods usually employing the use of alcohols and/or other chemical compounds such as polyethylene glycol (PEG). The resulting phage can be further purified using various density gradient/centrifugation methodologies. The resulting phage are then

15   chemically lysed, thereby releasing their nucleic acids that can be conveniently precipitated out of the supernatant to yield a viral nucleic acid supply of the phage of interest.

Exemplary bacteriophage are indicated in Table 1, along with sources where those phage may be obtained.

20   Exemplary bacteria include the reference bacteria for the identified viral strains, available from the same sources.

Characterizing Bacteriophage Genomes for ORFs

The third step involves systematically characterizing the genetic information

25   contained in the phage genome. Within this genetic information is the sequence of all RNAs and proteins encoded by the phage, including those that are essential or instrumental in inhibiting their host. This characterization is preferably done in a systematic fashion. For example, this can be done by first isolating high molecular weight genomic DNA from the phage using standard bacterial lysis methods, followed

30   by phage purification using density gradient ultracentrifugation, and extraction of nucleic acid from the purified phage preparation. The high molecular weight DNA is then analyzed to determine its size and to evaluate a proper strategy for its sequencing. The DNA is broken down into smaller size fragments by sonication or partial digestion with frequently cutting restriction enzymes such as Sau3A to yield

35   predominantly 1 to 2 kilobase length DNA, which DNA can then be resolved by gel electrophoresis followed by extraction from the gel.

The ends of the fragments are enzymatically treated to render them suitable for cloning and the pools of fragments are cloned in a bacterial plasmid to generate a

SD-63311.4

library of the phage genome. Several hundred of these random DNA fragments contained in the plasmid vector are isolated as clones after introduction into an appropriate bacterium, usually *Escherichia coli*. They are then individually expanded in culture and the DNA from each individual clone is purified. The nucleotide

5   sequences of the inserts of these clones are determined by standard automated or manual methods, using oligonucleotide primers located on either side of the cloning site to direct polymerase mediated sequencing (*e.g.*, the Sanger sequencing method or a modification of that method). Other sequencing methods can also be used.

The sequence of individual clones is then deposited in a computer, and

10   specific software programs (for example Sequencher™, Gene Codes Corp.) are used to look for overlap between the various sequences, resulting in ordering of contig sequences and ultimately providing the complete sequence of the entire bacteriophage genome (one such example is given in Table 2 for *Staphylococcus aureus* bacteriophage 77). This complete nucleotide sequence is preferably determined with

15   a redundancy of 3- to 5-fold (number of independent sequencing events covering the same region) in order to minimize sequencing errors.

Preferably, the bacterial strain used as a phage host should not possess any other innate plasmids, transposons, or other phage or incompatible sequences that would complicate or otherwise make the various manipulations and analyses more

20   difficult.

Commercially available computer software programs are used to translate the nucleotide sequence of the phage to identify all protein sequences encoded by the phage (hereafter called open reading frames or ORFs). As phages are known to transcribe their genome into RNA from both strands, in both directions, and

25   sometimes in more than one frame for the same sequence, this exercise is done for both strands and in all six possible reading frames. As evolutionary constraints have forced the phage to conserve all of its vital protein sequences in as small a genome as possible, it is straightforward to identify all the proteins encoded by the phage by simple examination of the 6 translation frames of the genome. Once these ORFs are

30   identified, they are cataloged into a phage proteome database (Table 3 lists ORFs identified from phage 77). This analysis is preferably performed for each phage under study. The process of ORF identification can be varied depending on the desired results. For example, the minimum length for the putative encoded polypeptide can be varied, and/or putative coding regions that have an associated Shine-Dalgarno

35   sequence can be selected. In the case of phage 77 ORFs, such parameter adjustment was performed and resulted in the identification of ORFs as listed herein. Different parameters had resulted in the identification of the ORFs listed in the preceding U.S.

Provisional Application 60/110,992, filed December 3, 1998, which is hereby incorporated by reference in its entirety.

Correlation of exemplary ORFs identified in that provisional application and as identified herein are shown in the following table:

ORF ID from 60/110,992	Genomic position	a.a. size	Start codon	ORF ID from 09/407,804	Genomic position	a.a. size	Start codon
77ORF016	2369-24024	251	TTG	77ORF017	23269-23982	237	ATG
77ORF019	39845-40501	218	ATA	77ORF019	39851-40501	216	ATG
77ORF050	29268-29564	98	ATG	77ORF182	29268-29564	98	ATG
77ORF050	29268-29564	98	ATG	77ORF043	29304-29564	86	ATG
77ORF067	34312-34551	79	CTG	77ORF104	34393-34551	52	ATG
77ORF146	29051-29212	53	ATG	77ORF102	29051-29212	53	ATG

Identifying and Characterizing Inhibitory Phage ORFs

The fourth step entails identifying the phage protein or proteins or RNA transcripts that have the ability to inhibit their bacterial hosts. This can be accomplished, for example, by either or both of two non-mutually exclusive methods. The first method makes use of bioinformatics. Over the past few years, a large amount of nucleotide sequence information and corresponding translated products have become available through large genome sequencing projects for a variety of organisms including mammals, insects, plants, unicellular eukaryotes (yeast and fungi), as well as several bacterial genomes such as *E. coli, Mycobacterium tuberculosis, Bacillus subtilis, Staphylococcus aureus* and many others. Such sequences have been deposited in public databases (for example, non-redundant sequence database at GenBank and SwissProt protein sequence database) (http://www.ncbi.nlm.nih.gov)) and can be freely accessed to compare any specific query sequence to those present in such databases. For example, GenBank contains over 1.6 billion nucleotides corresponding to 2.3 million sequence records. Several computer programs and servers (*e.g.*, TBLASTN) have been created to allow the rapid identification of homology between any given sequence from one organism to that of another present in such databases, and such programs are public and available free of charge.

In addition, it has been well established that basic biochemical pathways can be conserved in very distant organisms (for example bacteria and man), and that the proteins performing the various enzymatic steps in these pathways are themselves conserved at the amino acid sequence level. Thus, proteins performing similar functions (*e.g.* DNA repair, RNA transcription, RNA translation) have frequently preserved key structural signatures, identifiable by similarities across regions of

SD-63311.4

248/037

proteins (domains and motifs). The antimicrobials of the present invention will preferably target features and targets that are highly characteristic or conserved in microbes, and not higher organisms.

Most genomes encode individual proteins or groups of proteins that can be assembled into protein families that have been evolutionarily conserved. Therefore, similarity between a new query sequence and that of a member of a protein family (reference sequences from public databases) can immediately suggest a biochemical function for the novel query sequence, which in our case is a phage ORF.

The sequence homology between individual members of evolutionarily distant members of a protein family is usually not randomly distributed along the entire length of the sequence but is often clustered into "motifs". These correspond to key three-dimensional folds that form key catalytic and/or regulatory structures that perform key biochemical function(s) for the group of proteins. Commercially available computer software programs can identify such motifs in a new query sequence, again providing functional information for the query sequence. Such structural and functional motifs have also been derived from the combined analysis of primary sequence databases (protein sequences) and protein structure databases (X-ray crystallography, nuclear magnetic resonance) using so-called "threading" methods (Rost B,l and Sander C. (1996).*Ann. Rev. Biophy. Biomol. Struct.* 25, 113-136).

Such motifs and folds are themselves deposited in public databases which can be directly accessed (for example, SwissProt database; 3D-ALI at EMBL, Heidelberg; PROSITE). This basic exercise leads to a structural homology map in which each of the phage ORFs has been probed for such similarities, and where initial structural and functional hits are identified (selected examples of sequence homologies detected between individual ORFs from the genome of *Staphylococcus aureus* bacteriophage 77 and sequences deposited in public databases are shown in Table 5; listed are the proteins showing homologies and the TBLASTN scores quantifying the degree of sequence similarity between the two compared sequences).

This analysis can point out phage proteins with similarity to proteins from other phages (such as those for *E. coli*) playing an important role in the basic biochemical pathways of the phage (such as DNA replication, RNA transcription, tRNAs, coat protein and assembly). Selected examples of such proteins are shown in Tble 5. Therefore, this analysis enables identification and elimination of non-essential ORFs as candidates for an inhibitor function, as well as the identification of (potentially) useful ones.

In addition, this analysis can point out specific ORFs as possible inhibitor ORFs. For example these ORFs may encode proteins or enzymes that alter bacterial cell structure, metabolism or physiology, and ultimately viability. Examples of such

SD-63311.4

proteins present in the genome of *Staphylococcus aureus* bacteriophage 77 include orf14 (deoxyuridine triphosphatase from bacteriophage T5), and orf15 (sialidase).

5   In addition, it is well known that bacterial and eukaryotic viruses can usurp pathways from their host in order to use them to their advantage in blocking host cellular pathways upon infection. The phage can achieve this, for example, by overexpressing part or whole host-related sequences which are themselves regulating or rate limiting in key biochemical pathways of the host. The identification of sequence similarity between phage ORFs and bacterial host genome sequences will be highly indicative of such a mechanism (Selected examples of such homologies are

10   listed in Table 5, e.g. orf4 (homologous to autolysin), orf20 (hypothetical protein from *Staphyloccus aureus*) and orf29 (hypothetical protein from *Staphyloccus aureus*). These ORFs can be analyzed by a standard biochemical approach to directly test their inhibitor functions (*e.g.*, as described below).

Alternatively, a homology search may reveal that a given phage ORF is related

15   to a protein present in the databases having an activity known to be inhibitory, (*e.g.* inhibitor of host RNA polymerase by *E. coli* bacteriophage T7. Such a finding would implicate the phage ORF product in a related activity. This will also suggest that a new antimicrobial could be derived by a mimetic approach (*e.g.*, peptidomimetic) imitating this function or by a small molecule inhibitor to the bacterial target of the

20   phage ORF, or any steps in the relevant host metabolic pathway, *e.g.*, high throughput screening of small molecule libraries. Selected examples of such similarity between ORFs of *Staphyloccus aureus* bacteriophage 77 and proteins with inhibitor functions for bacterial hosts are listed in Table 5. These include orf9 (similar to bacteriophage P1 *kilA* function), and orf4 (autolysin of *Staphylococcus aureus*, amidase enzymatic

25   activity).

A reason for the biochemical study of individual ORFs for inhibitor function is that their expression or overexpression will block cellular pathways of the host, ultimately leading to arrest and/or inhibition of host metabolism. In addition, such ORFs can alter host metabolism in different ways, including modification of

30   pathogenicity. Therefore, individual ORFs identified above are expressed, preferably overexpressed, in the host and the effect of this expression or overexpression on host metabolism and viability is measured. This approach can be systematically applied to every ORF of the phage, if necessary, and does not rely on the absolute identification of candidate ORFs by bioinformatics. Individual ORFs are resynthesized from the

35   phage genomic DNA, *e.g.*, by the polymerase chain reaction (PCR), preferably using oligonucleotide primers flanking the ORF on either side. These single ORFs are preferably engineered so that they contain appropriate cloning sites at their extremities to allow their introduction into a new bacterial expression plasmid, allowing

propagation in a standard bacterial host such as *E. coli,* but containing the necessary information for plasmid replication in the target microbe such as *S. aureus* (hereafter referred to as shuttle vector). Shuttle vectors and their use are well known in the art.

5   Such shuttle vectors preferably also contain regulatory sequences that allow inducible expression of the introduced ORF. As the candidate ORF may encode an inhibitor function that will eliminate the host, it is beneficial that it not be expressed prior to testing for activity. Thus, screening for such sequences when expressed in a constitutive fashion is less likely to be successful when the inhibitor is lethal. In the exemplary inducible system presented in Figures 1A, 1B, and 2, regulatory sequences

10 from the *ars* operon of *S. aureus* are used to direct individual ORF expression in *S. aureus.* The *ars* operon encodes a series of proteins which normally mediate the extrusion of arsenite and other trivalent oxyanions from the cells when they are exposed to such toxic substances in their environment. The operon encoding this detoxifying mechanism is normally silent and only induced when arsenite-related

15 compounds are present. (Tauriainen, S. et al. (1997) *App. Env. Microb.,* Vol. 63, No. 11, p. 4456-4461.)

Therefore, individual phage ORFs can be expressed in *S. aureus* in an inducible fashion by adding to the culture medium non-toxic arsenite concentrations during the growth of individual *S. aureus* clones expressing such individual phage

20 ORFs. Toxicity of the phage inhibitor ORF for the host is monitored by reduction or arrest of growth under induction conditions, as measured by optical density in liquid culture or after plating the induced cultures on solid medium. Subsequently, interference of the phage ORF with the host biochemical pathways ultimately leading to reduced or arrested host metabolism can be measured by pulse-chase experiments

25 using radiolabeled precursors of either DNA replication, RNA transcription, or protein synthesis.

Those skilled in the art are familiar with a variety of other inducible systems which can also be used for the controlled expression of phage ORFs, including, for example, lactose (see *e.g.,* Stratagene's LacSwitch™II system; La Jolla, CA) and

30 tetracycline-based systems (see, *e.g.* Clontech's Tet On/Tet Off™ system; Palo Alto, CA). The arsenite-inducible system described is further depicted in Figures 1A, 1B, and 2.

The selection or construction of shuttle vectors and the selection and use of inducible systems are well known and thus other shuttle vectors appropriate for other

35 bacteria can be readily provided by those skilled in the art.

Standard methodologies for expressing proteins from constructs, and isolating and manipulating those proteins, for example in cross-linking and affinity chromatography studies, may be found in various commonly available and known

SD-63311.4

laboratory manuals. See, *e.g.*, Current Protocols in Protein Science, John Wiley & Sons, Secaucus, N.J., and Maniatis, T. et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor University Press, Cold Spring, N.Y.

It has been found that certain phage or other viruses inhibit host cells, at least in part, by producing an antisense RNA which binds to and inhibits translation from a bacterial RNA seqeunce. Thus, in the case of potentially inhibitor RNA transcripts encoded by the phage genome, a strong indicator of a possible inhibitory function is provided by the identification of phage sequence which is the identical to or fully complementary (or with only a small percentage of mismatch, *e.g.*, <10%, preferably less than 5%, most preferably less than 3%, to a bacterial sequence. This approach is convenient in the case of bacteria which have been essentially completely sequenced, as the comparison can be performed by computer using public database information.

The inhibitory effect of the transcript can be confirmed using expression of the phage sequence in a host bacterium. If needed, such inhibitory can also be tested by transfecting the cells with a vector which will transcribe the phage sequence to form RNA in such manner that the RNA produced will not be translated into a polypeptide. Inhibition under such conditions provides a strong indication that the inhibition is due to the transcript rather than to an encoded polypeptide.

In an alternative, the expression of an ORF in a host bacterium is found to be inhibitory, but the inhibition if found to be due to an RNA product of the genomic coding region. For antisense inhibition, the sequence of the bacterial target nucleic acid sequence can be identified by inspection of the phage sequence, and the full sequence of the relevant coding region for the bacterial product can be found from a database of the bacterial genomic sequence or can be isolated by standard techniques (*e.g.*, a clone in a genomic library can be isolated which contains the full bacterial ORF, and then sequenced).

In either case, the identification of a target which is inhibited by an RNA transcript produced by a phage provides both the possible inhibition of bacteria naturally containing the same target nucleic acid sequence, as well as the ability to use the target sequence in screening for other types of compounds which will act directly on the target nucleic acid sequence or on a polypeptide product expressed or regulated, at least in part, by the target of the inhibitory phage RNA.

In some cases it will be found that the target of an inhibitory phage RNA or protein has previously been found to be a target of an inhibitory phage RNA or protein has previously been found to be a target for an antibacterial agent. In such cases, the phage inhibitor can still provide useful information if it is found that the phage-encoded product acts at a different site than the previously identified antibacterial agent or inhibitor, *i.e.*, acts at a phage-specific site. For many targets,

SD-63311.4

action at a different site provides highly beneficial characteristics and/or information. For example, an alternate site of inhibitor action can at least partially overcome a resistance mechanism in a bacterium. As an illustration, in many cases, resistance is due, in large part, to altered binding characteristics of the immediate target to the

5   antibacterial agent. The altered binding is due to a structural change which prevents or destabilizes the binding. However, the structural change is frequently quite local, so that compounds which bind at different local sites will b unaffected or affected to a much lesser degree. Indeed, in some cases the local sites will be on a different molecule and so may be completely unaffected by the local structural change creating

10   resistance to the original agent(s). An example of resistance due to altered binding is provided by methicillin-resistant *Staphylococcus aureus*, in which the resistance is due to an altered penicillin-binding protein.

In other cases, a new site of action can have improved accessibility as compared to a site acted on by a previously identified agent. This can, for example,

15   assist in allowing effective treatment at lower doses, or in allowing access by a larger range of types of compounds, potentially allowing identification of more potential active agents.

Another advantage is that the structural characteristics of a different site of action will lead to identification and/or development of inhibitors with different

20   structures and different pharmacological parameter. This can allow a greater range of possibilities when selecting an antibacterial agent.

Yet further, different sites often produce different inhibitory characteristics in the target organism. This is commonly the case for multi-domain target proteins. Thus, inhibition targeting an alternate site can produce more efficacious action, *e.g.*,

25   faster killing, slower development of resistance, lower numbers of surviving cells, and different secondary effects (for example, different nutrient utilization).

Validating Identified Inhibitory Phage ORFs

A fifth step involves validating the identified phage inhibitor ORF by

30   independent methods, and delineating further possible smaller segments of the ORFs that have inhibitory activity. Several methods exist to validate the role of the identified ORF as an inhibitor ORF.

One example utilizes the creation of a mutant variant of the phage ORF in which the candidate ORF carries a partial or complete loss-of-function mutation that

35   is measurable as compared with the non-mutant ORF. Comparison of the effects of expression of the loss of function mutant with the normal ORF provides confirmation of the identification of an inhibitor ORF where the loss of function mutant provides a

measurably lower level of inhibition, preferably no inhibition. The loss of function may be conditional, *e.g.*, temperature sensitive.

Once validation of the inhibitor ORF is achieved, a bi-directional deletion analysis can be carried out using the same experimental system to identify the

5   minimal polypeptide segment that has inhibitor activity. This may be carried out by a variety of means, *e.g.*, by exonuclease or PCR methodologies, and is used to determine if a relatively small segment of the ORF (*i.e.*, the product of the ORF) still possesses inhibitory activity when isolated away from its native sequence. If so, a portion of the ORF encoding this "active portion" can be used as a template for the

10 synthesis of novel anti-microbial agents and further allowing derivation of the peptide sequence, *e.g.*, using modified peptides and/or peptidomimetics.

In creation of certain peptidomimetics, the peptide backbone is transformed into a carbon-based hydrophobic structure that can retain inhibitor activity against the bacterium. This is done by standard medicinal chemistry methods, typically

15 monitored by measuring growth inhibition of the various molecules in liquid cultures or on solid medium. These mimetics can also represent lead compounds for the development of novel antibiotics.

Recently, a major effort has been undertaken by the pharmaceutical industry and their biotechnology partners for the sequencing of bacterial pathogen genomes.

20 The rationale is that the systematic sequencing of the genome will identify all of the bacterial proteins and therefore this proteome will be the target for designing novel inhibitor antibiotics. Although systematic, this approach has several major problems. The first is that analysis of primary amino acid sequences of bacterial proteins does not immediately reveal which protein will be essential for viability of the bacterium,

25 and target validation is thus a major issue. The second problem is one of redundancy, as several biochemical pathways are either structurally duplicated in bacteria (different isoforms of the same enzyme), or functionally duplicated by the presence of salvage pathways in the event of a metabolic block in one pathway (different nutritional conditions). The third is that even a valid target may not be structurally or

30 functionally amenable to inhibition by small molecules because of inaccessibility (sequestration of target).

Therefore, there is considerable interest within the pharmaceutical and biotechnology industry in identifying key targets for drug discovery amongst the mass of novel targets generated by large-scale genomic sequencing projects.

35 On the other hand, and underscoring the instant invention, the phages herein described have, over millions of years, evolved specific mechanisms to target such key biochemical pathways and proteins. In the few cases where inhibition by phages has been elucidated (*e.g.*, see ref. 3), such bacterial targets are invariably rate-limiting

in their respective biochemical pathways, are not redundant, and/or are readily accessible for inhibition by the phage (or by another inhibitory compound). Therefore, the sixth step of this invention involves identifying the host biochemical pathways and proteins that are targeted by the phage inhibitory mechanisms.

5

Identifying, Validating, and Characterizing Bacterial Host Target Proteins and
Affected Pathways

A rationale for this step is that the inhibitor ORF product from the phage physically interacts with and/or modifies certain microbial host components to block
10   their function. Exemplary approaches which can be used to identify the host bacterial pathways and proteins that interact with, and preferably also are inhibited by, phage ORF product(s) are described below.

The first approach is a genetic screen to determine physiological protein:protein interaction, for example, using a yeast two hybrid system. In this
15   assay, the phage ORF is fused to the carboxyl terminus of the yeast Gal4 activation domain II (amino acids 768-881) to create a bait vector. A cDNA library of cloned *S. aureus* sequences which have been engineered into a plasmid where the *S. aureus* sequences are fused to the DNA binding domain of Gal4 is also generated. These plasmids are introduced alone, or in combination, into yeast strain Y190 - previously
20   engineered with chromosomally integrated copies of the *E. coli lacZ* and the selectable HIS3 genes, both under Gal4 regulation (Durfee, T., Becherer, K., Chen, P.-L., Yeh, S.-H., Yang, Y., Kilburn, A.E., Lee, W.-H., and Elledge, S.J. (1993). *Genes & Dev.* 7, 555-569). If the two proteins expressed in yeast interact, the resulting complex will activate transcription from promoters containing Gal4 binding sites. A
25   *lacZ* and His3 gene, each driven by a promoter containing Gal4 binding sites, have been integrated into the genome of the host yeast system used for measuring protein-protein interactions. Such a system provides a physiological environment in which to detect potential protein interactions. This system has been extensively used to identify novel protein-protein interaction partners and to map the sites required for interaction
30   (for example, to identify interacting partners of translation factors (Qiu, H., Garcia-Barrio, M.T., and Hinnebusch, A.G. (1998). *Mol & Cell Biology* 18, 2697-2711), transcription factors (Katagiri, T., Saito, H., Shinohara, A., Ogawa, H., Kamada, N., Nakamura ,Y., and Miki, Y. (1998). Genes, *Chromosomes & Cancer* 21, 217-222), and proteins involved in signal transduction (Endo, T.A., Masuhara, M., Yokouchi,

SD-63311.4

M., Suzuki, R., Sakamoto, H., Mitsui, K., Matsumoto, A., Tanimura, S., Ohtsubo, M., Misawa, H., Miyazaki, T., Leonor N., Taniguchi, T., Fujita, T., Kanakura, Y., Komiya, S., and Yoshimura, A. *Nature.* 387, 921-924). This approach has also been used in many published reports to identify interaction between mammalian viral and

5   mammalian cell proteins.

For example, the non-structural protein NS1 of parvovirus is essential for viral DNA amplification and gene expression and is also the major cytopathic effector of these viruses. A yeast two-hybrid screen with NS1 identified a novel cellular protein of unknown function that interacts with NS-1, called SGT, for small glutamine-rich

10   tetratricopeptide repeat (TPR)-containing protein (Cziepluch C. Kordes E. Poirey R. Grewenig A. Rommelaere, J, and Jauniaux JC. (1998) *J Virol.* 72, 4149-4156). In another screen, the adenovirus E3 protein was recently shown to interact with a novel tumor necrosis factor alpha-inducible protein and to modulate some of the activities of E3 (Li Y. Kang J. and Horwitz M.S. (1998). *Mol & Cell Biol.* 18, 1601-1610). In yet

15   another recent screen, the herpes simplex virus 1 alpha regulatory protein ICP0 was found to interact with (and stabilize) the cell cycle regulator cyclin D3 (Kawaguchi Y. Van Sant C. and Roizman B. (1997). *J Virol.* 71,7328-7336).

Another two-hybrid system for identifying protein:protein interactions is commercially available from STRATEGENE™ as the CYTO-TRAP™ system

20   (Chang et al., *Strategies Newsletter* 11(3), 65-68 (1998)(from Stratagene)). The system is a yeast-based method for detecting protein:protein interactions *in vivo*, using activation of the Ras signal transduction cascade by localizing a signal pathway component, human Sos (hSos), to its activation site in the yeast plasma membrane. The system uses a temperature-sensitive *Saccharomyces cerevisiae* mutant, strain

25   cdc25H, which contains a point mutation at amino acid residue 1328 of the cdc25 gene. This gene encodes a guanyl nucleotide exchange factor which binds and activates Ras, leading to cell growth. The mutation in the cdc25 gene prevents host growth at 37°C, but at a permissive temperature of 25°C, growth is normal. The system utilizes the ability of (hSos) to complement the cdc25 defect and activate the

30   yeast Ras signaling pathway. Once (hSos) is expressed and localized to the plasma membrane, the cdc25H yeast strain grows at 37°C. Localizing hSos to the plasma membrane occurs through a protein:protein interaction. A protein of interest, or bait, is expressed as a fusion protein with hSos. The library, or target proteins are

expressed with the myristylation membrane-localization signal. The yeast cells are then incubated under restrictive conditions (37°C). If the bait and the target protein interact, the hSos protein is recruited to the membrane, activating the Ras signaling pathway and allowing the cdc25H yeast strain to grow at the restrictive temperature.

5   The second approach is based on identifying protein:protein interactions between the phage ORF product and bacterial *S. aureus,* e.g., proteins using a biochemical approach based, for example, on affinity chromatography. This approach has been used, for example, to identify interactions between lambda phage proteins and proteins from their *E. coli* host (Sopta, M., Carthew, R.W., and Greenblatt, J.

10   (1985) *J. Biol. Chem.* 260, 10353-10369). The phage ORF is fused to a peptide tag (*e.g.* glutathione-S-transferase ("GST"), 6xHIS, ("HIS") and/or calmodulin binding protein ("CPB")) within a commercially available plasmid vector that directs high level expression on induction of a suitably responsive promoter driving the fusion's expression. The translated fusion protein is expressed in *E. coli*, purified, and

15   immobilized on a solid phase matrix via, for example the tag. Total cell extracts from the host bacterium, *e.g.*, *S. aureus*, are then passed through the affinity matrix containing the immobilized phage ORF fusion protein; host proteins retained on the column are then eluted under different conditions of ionic strength, pH, detergents etc., and characterized by gel electrophoresis and other techniques. Appropriate

20   controls are run to guard against nonspecific binding to the resin. Target proteins thus recovered should be enriched for the phage protein/peptide of interest and are subsequently electrophoretically or otherwise separated, purified, sequenced, or biochemically analyzed. Usually sequencing entails individual digestion of the proteins to completion with a protease (*e.g.*-trypsin), followed by molecular mass and

25   amino acid composition and sequence determination using, for example, mass spectrometry, *e.g.*, by MALDI-TOF technology (Qin, J., Fenyo, D., Zhao, Y., Hall, W.W., Chao, D.M., Wilson, C.J., Young, R.A. and Chait, B.T. (1997). *Anal. Chem.* 69, 3995-4001).

           The sequence of the individual peptides from a single protein are then

30   analyzed by the bioinformatics approach described above to identify the *S. aureus* protein interacting with the phage ORF. This analysis is performed by a computer search of the *S. aureus* genome for an identified sequence. Alternatively, all tryptic peptide fragments of the *S. aureus* genome can be predicted by computer software,

and the molecular mass of such fragments compared to the molecular mass of the peptides obtained from each interacting protein eluted from the affinity matrix. The responsible gene sequence can be obtained, for example by using synthetic degenerate nucleic acid sequences to pull out the corresponding homologous bacterial sequence.

5   Alternatively, antibodies can be generated against the peptide and used to isolate nascent peptide/mRNA transcript complexes, from which the mRNA can be reverse transcribed, cloned, and further characterized using the procedures discussed herein.

A variety of other binding assay methods are known in the art and can be used to identify interactions between phage proteins and bacterial proteins or other bacterial

10   cell components. Such methods which allow or provide identification of the bacterial component can be used in this invention for identifying putative targets.

Validation of the interaction between the phage ORF product and the bacterial proteins or other components can be obtained by a second independent assay (*e.g.,* co-immunoprecipitation or protein-protein crosslinking experiments (Qiu, H., Garcia-

15   Barrio, M.T., and Hinnebusch, A.G. (1998). *Mol & Cell Biology* 18, 2697-2711; Brown, S. and Blumenthal, T. (1976). *Proc. Natl. Acad. Sci. USA* 73, 1131-1135)).

Finally, the essential nature of the identified bacterial proteins is preferably determined genetically by creating a constitutive or inducible partial or complete loss-of-function mutation in the gene encoding the identified interacting bacterial protein.

20   This mutant is then tested for bacterial survival and replication.

The protein target of the phage inhibitor function can also be identified using a genetic approach. Two exemplary approaches will be delineated here. The first approach involves the overexpression of a predetermined phage inhibitor protein in mutagenized host bacteria, *e.g., S. aureus,* followed by plating the cells and searching

25   for colonies that can survive the inhibitor. These colonies will then be grown, their DNA extracted and cloned into an expression vector that contains a replicon of a different incompatibility group, and preferably having a different selectable marker than the plasmid expressing the phage inhibitor. Thus, host DNA fragments from the mutant that can protect the cell from phage ORF inhibition can be sequenced and

30   compared with that of the bacterial host to determine in which gene the mutation lies. This approach allows rapid determination of the targets and pathways that are affected by the inhibitor.

Alternatively, the bacterial targets can be determined in the absence of selecting for mutations using an approach known as "multicopy suppression". In this approach, the DNA from the wild type host is cloned into an expression vector that can coexist, as previously described, with one containing a predetermined phage

5   inhibitor. Those plasmids that contain host DNA fragments and genes that protect the host from the phage inhibitor can then be isolated and sequenced to identify putative targets and pathways in the host bacteria.

Regardless of the specific mode of identification, screening assays may additionally utilize gene fusions to specific "reporter genes" to identify a bacterial

10   gene(s) whose expression is affected when the host target pathway is affected by the phage inhibitor. Such gene fusions can be used to search a number of small molecule compounds for inhibitors that may affect this pathway and thus cause cell inhibition. This approach will allow the screening of a large number of molecules on petri dishes or 96-well format by monitoring for a simple color change in the bacterial colonies.

15   In this manner, we can validate host targets and classes of compounds for further study and clinical development. These inhibitors also represent lead compounds for the development of other antibiotics.

Bioinformatics and comparative genomics are preferably then applied to the identified bacterial gene products to predict biochemical function. The biochemical

20   activity of the protein can be verified *in vitro* in cell free assays or *in vivo* in intact cells. *In vitro* biochemical assays utilizing cell-free extracts or purified protein are established as a basis for the screening and development of inhibitors.

These inhibitors, preferably small molecule inhibitors, may comprise peptides, antibodies, products from natural sources such as fungal or plant extracts or small

25   molecule organic compounds. In general, small molecule organic compounds are preferred. These compounds may, for example, be identified within large compound libraries, including combinatorial libraries. For example, a plurality of compounds, preferably a large number of compounds can be screened to determine whether any of the compounds binds or otherwise disrupts or inhibits the identified bacterial target.

30   Compounds identified as having any of these activities can then be evaluated further in cell culture and/or animal model systems to determine the pharmacological properties of the compound, including the specific anti-microbial ability of the compound.

SD-63311.4

For mixtures of natural products, including crude preparations, once a preparation or fraction of a preparation is shown the have an anti-microbial activity, the active substance can be isolated and identified using techniques well known in the art, if the compound is not already available in a purified form.

5   Identified compounds possessing anti-microbial activity and similar compounds having structural similarity can be further evaluated and, if necessary, derivatized according to synthesis and/or modification methods available in the art selected as appropriate for the particular starting molecule.

10   Derivatization of identified anti-microbials

In cases where the identified anti-microbials above might represent peptidal compunds, the *in vivo* effectiveness of such compounds may be advantageously enhanced by chemical modification using the natural polypeptide as a starting point and incorporating changes that provide advantages for use, for example, increased stability to proteolytic degradation, reduced antigenicity, improved tissue penetration, and/or improved delivery characteristics.

In addition to active modifications and derivative creations, it can also be useful to provide inactive modifications or derivatives for use as negative controls or introduction of immunologic tolerance. For example, a biologically inactive derivative which has essentially the same epitopes as the corresponding natural antimicrobial can be used to induce immunological tolerance in a patient being treated. The induction of tolerance can then allow uninterrupted treatment with the active anti-microbial to continue for a significantly longer period of time.

25   Modified anti-microbial polypeptides and derivatives can be produced using a number of different types of modifications to the amino acid chain. Many such methods are known to those skilled in the art. The changes can include, for example, reduction of the size of the molecule, and/or the modification of the amino acid sequence of the molecule. In addition, a variety of different chemical modifications of the naturally occurring polypeptide can be used, either with or without modifications to the amino acid sequence or size of the molecule. Such chemical modifications can, for example, include the incorporation of modified or non-natural amino acids or non-amino acid moieties during synthesis of the peptide chain, or the post-synthesis modification of incorporated chain moieties.

SD-63311.4

The oligopeptides of this invention can be synthesized chemically or through an appropriate gene expression system. Synthetic peptides can include both naturally occurring amino acids and laboratory synthesized, modified amino acids.

5   Also provided herein are functional derivatives of anti-microbial proteins or polypeptides. By "functional derivative" is meant a "chemical derivative," "fragment," "variant," "chimera," or "hybrid" of the polypeptide or protein, which terms are defined below. A functional derivative retains at least a portion of the function of the protein, for example reactivity with a specific antibody, enzymatic activity or binding activity.

10   A "chemical derivative" of the complex contains additional chemical moieties not normally a part of the protein or peptide. Such moieties may improve the molecule's solubility, absorption, biological half-life, and the like. The moieties may alternatively decrease the toxicity of the molecule, eliminate or attenuate any undesirable side effect of the molecule, and the like. Moieties capable of mediating

15   such effects are disclosed in Alfonso and Gennaro (1995). Procedures for coupling such moieties to a molecule are well known in the art. Covalent modifications of the protein or peptides are included within the scope of this invention. Such modifications may be introduced into the molecule by reacting targeted amino acid residues of the peptide with an organic derivatizing agent that is capable of reacting

20   with selected side chains or terminal residues, as described below.

Cysteinyl residues most commonly are reacted with alpha-haloacetates (and corresponding amines), such as chloroacetic acid or chloroacetamide, to give carboxymethyl or carboxyamidomethyl derivatives. Cysteinyl residues also are derivatized by reaction with bromotrifluoroacetone, chloroacetyl phosphate, N-

25   alkylmaleimides, 3-nitro-2-pyridyl disulfide, methyl 2-pyridyl disulfide, p-chloro-mercuribenzoate, 2-chloromercuri-4-nitrophenol, or chloro-7-nitrobenzo-2-oxa-1,3-diazole.

Histidyl residues are derivatized by reaction with diethylprocarbonate at pH 5.5-7.0 because this agent is relatively specific for the histidyl side chain. Para-

30   bromophenacyl bromide also is useful; the reaction is preferably performed in 0.1 M sodium cacodylate at pH 6.0.

Lysinyl and amino terminal residues are reacted with succinic or other carboxylic acid anhydrides. Derivatization with these agents has the effect of

SD-63311.4

reversing the charge of the lysinyl residues. Other suitable reagents for derivatizing
primary amine- containing residues include imidoesters such as methyl
picolinimidate; pyridoxal phosphate; pyridoxal; chloroborohydride;
trinitrobenzenesulfonic acid; O-methylisourea; 2,4 pentanedione; and transaminase-

5   catalyzed reaction with glyoxylate.

        Arginyl residues are modified by reaction with one or several conventional
reagents, among them phenylglyoxal, 2,3-butanedione, 1,2-cyclohexanedione, and
ninhydrin. Derivatization of arginine residues requires that the reaction be performed
in alkaline conditions because of the high pK_a of the guanidine functional group.

10   Furthermore, these reagents may react with the groups of lysine as well as the arginine
alpha-amino group.

        Tyrosyl residues are well-known targets of modification for introduction of
spectral labels by reaction with aromatic diazonium compounds or tetranitromethane.
Most commonly, N-acetylimidizol and tetranitromethane are used to form O-acetyl

15   tyrosyl species and 3-nitro derivatives, respectively.

        Carboxyl side groups (aspartyl or glutamyl) are selectively modified by
reaction carbodiimide (R'-N-C-N-R') such as 1-cyclohexyl-3-(2-morpholinyl(4-ethyl)
carbodiimide or 1-ethyl-3-(4-azonia-4,4-dimethylpentyl) carbodiimide. Furthermore,
aspartyl and glutamyl residues are converted to asparaginyl and glutaminyl residues

20   by reaction with ammonium ions.

        Glutaminyl and asparaginyl residues are frequently deamidated to the
corresponding glutamyl and aspartyl residues. Alternatively, these residues are
deamidated under mildly acidic conditions. Either form of these residues falls within
the scope of this invention.

25   Derivatization with bifunctional agents is useful, for example, for cross-
linking component peptides to each other or the complex to a water-insoluble support
matrix or to other macromolecular carriers. Commonly used cross-linking agents
include, for example, 1,1-bis (diazoacetyl)-2-phenylethane, glutaraldehyde, N-
hydroxysuccinimide esters, for example, esters with 4-azidosalicylic acid, homobi-

30   functional imidoesters, including disuccinimidyl esters such as 3,3'-
dithiobis(succinimidylpropionate), and bifunctional maleimides such as bis-N-
maleimido-1,8-octane. Derivatizing agents such as methyl-3-[p-azidophenyl)
dithiolpropioimidate yield photoactivatable intermediates that are capable of forming

crosslinks in the presence of light. Alternatively, reactive water-insoluble matrices such as cyanogen bromide-activated carbohydrates and the reactive substrates described in U.S. Patent Nos. 3,969,287; 3,691,016; 4,195,128; 4,247,642; 4,229,537; and 4,330,440 are employed for protein immobilization.

5   Other modifications include hydroxylation of proline and lysine, phosphorylation of hydroxyl groups of seryl or threonyl residues, methylation of the alpha-amino groups of lysine, arginine, and histidine side chains (Creighton, T.E., *Proteins: Structure and Molecular Properties*, W.H. Freeman & Co., San Francisco, pp. 79-86 (1983)), acetylation of the N-terminal amine, and, in some instances,

10   amidation of the C-terminal carboxyl groups.

          Such derivatized moieties may improve the stability, solubility, absorption, biological half life, and the like. The moieties may alternatively eliminate or attenuate any undesirable side effect of the protein complex. Moieties capable of mediating such effects are disclosed, for example, in Alfonso and Gennaro (1995).

15   The term "fragment" is used to indicate a polypeptide derived from the amino acid sequence of the protein or polypeptide having a length less than the full-length polypeptide from which it has been derived. Such a fragment may, for example, be produced by proteolytic cleavage of the full-length protein. Preferably, the fragment is obtained recombinantly by appropriately modifying the DNA sequence encoding

20   the proteins to delete one or more amino acids at one or more sites of the C-terminus, N-terminus, and/or within the native sequence.

          Another functional derivative intended to be within the scope of the present invention is a "variant" polypeptide which either lacks one or more amino acids or contains additional or substituted amino acids relative to the native polypeptide. The

25   variant may be derived from a naturally occurring polypeptide by appropriately modifying the protein DNA coding sequence to add, remove, and/or to modify codons for one or more amino acids at one or more sites of the C-terminus, N-terminus, and/or within the native sequence.

          A functional derivative of a protein or polypeptide with deleted, inserted

30   and/or substituted amino acid residues may be prepared using standard techniques well-known to those of ordinary skill in the art. For example, the modified components of the functional derivatives may be produced using site-directed mutagenesis techniques (as exemplified by Adelman et al., 1983, *DNA* 2:183;

Sambrook et al., 1989) wherein nucleotides in the DNA coding sequence are modified such that a modified coding sequence is produced, and thereafter expressing this recombinant DNA in a prokaryotic or eukaryotic host cell, using techniques such as those described above. Alternatively, components of functional derivatives of

5   complexes with amino acid deletions, insertions and/or substitutions may be conveniently prepared by direct chemical synthesis, using methods well-known in the art.

Insofar as other anti-microbial inhibitor compounds identified by the invention described herein may not be peptidal in nature, other chemical techniques exist to

10   allow their suitable modification, as well, and according the desirable principles discussed above.

Administration and Pharmaceutical Compositions

15   For the therapeutic and prophylactic treatment of infection, the preferred method of preparation or administration of anti-microbial compounds will generally vary depending on the precise identity and nature of the anti-microbial being delivered. Thus, those skilled in the art will understand that administration methods known in the art will also be appropriate for the compounds of this invention.

20   The particularly desired anti-microbial can be administered to a patient either by itself, or in pharmaceutical compositions where it is mixed with suitable carriers or excipient(s). In treating an infection, a therapeutically effective amount of an agent or agents is administered. A therapeutically effective dose refers to that amount of the compound that results in amelioration of one or more symptoms of bacterial infection

25   and/or a prolongation of patient survival or patient comfort.

Toxicity, therapeutic and prophylactic efficacy of anti-microbials can be determined by standard pharmaceutical procedures in cell cultures and/or experimental organisms such as animals, e.g., for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in

30   50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD_{50}/ED_{50}. Compounds which exhibit large therapeutic indices are preferred. The data obtained from these cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of

circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized.

5   For any compound identified and used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. Such information can be used to more accurately determine useful doses in organisms such as plants and animals, preferably mammals, and most preferably humans. Levels in plasma may be measured, for example, by HPLC or other means appropriate for detection of the particular compound.

10   The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition (see *e.g.* Fingl et. al., in The Pharmacological Basis of Therapeutics, 1975, Ch. 1 p.1).

It should be noted that the attending physician would know how and when to terminate, interrupt, or adjust administration due to toxicity, organ dysfunction, or

15   other systemic malady. Conversely, the attending physician would also know to adjust treatment to higher levels if the clinical response were not adequate (precluding toxicity). The magnitude of an administered dose in the management of the disorder of interest will vary with the severity of the condition to be treated and the route of administration. The severity of the condition may, for example, be evaluated, in part,

20   by standard prognostic evaluation methods. Further, the dose and perhaps dose frequency, will also vary according to the age, body weight, and response of the individual patient. A program comparable to that discussed above also may be used in veterinary or phyto medicine.

Depending on the specific infection target being treated and the method

25   selected, such agents may be formulated and administered systemically or locally, i.e., topically. Techniques for formulation and administration may be found in Alfonso and Gennaro (1995). Suitable routes may include , for example, oral, rectal, transdermal, vaginal, transmucosal, intestinal, parenteral, intramuscular, subcutaneous, or intramedullary injections, as well as intrathecal, intravenous, or

30   intraperitoneal injections.

For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiological saline buffer. For transmucosal administration,

penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

Use of pharmaceutically acceptable carriers to formulate identified anti-microbials of the present invention into dosages suitable for systemic administration is within the scope of the invention. With proper choice of carrier and suitable manufacturing practice, the compositions of the present invention, in particular those formulated as solutions, may be administered parenterally, such as by intravenous injection. Appropriate compounds can be formulated readily using pharmaceutically acceptable carriers well known in the art into dosages suitable for oral administration. Such carriers enable the compounds of the invention to be formulated as tablets, pills, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated.

Agents intended to be administered intracellularly may be administered using techniques well known to those of ordinary skill in the art. For example, such agents may be encapsulated into liposomes, then administered as described above. Liposomes are spherical lipid bilayers with aqueous interiors. All molecules present in an aqueous solution at the time of liposome formation are incorporated into the aqueous interior. The liposomal contents are both protected from the external microenvironment and, because liposomes fuse with cell membranes, are efficiently delivered into the cell cytoplasm. Additionally, due to their hydrophobicity, small organic molecules may be directly administered intracellularly.

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. Determination of the effective amounts is well within the capability of those skilled in the art.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. The preparations formulated for oral administration may be in the form of tablets, dragees, capsules, or solutions, including those formulated for delayed release or only to be released when the pharmaceutical reaches the small or large intestine.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is itself known, *e.g.*, by means of conventional mixing, dissolving, granulating, dragee-making, levitating, emulsifying, encapsulating, entrapping or lyophilizing processes.

5   Pharmaceutical formulations for parenteral administration include aqueous solutions of the active anti-microbial compounds in water-soluble form. Alternatively, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

15   Pharmaceutical preparations for oral use can be obtained by combining the active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate.

25   Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active

ingredients in admixture with filler such as lactose, binders such as starches, and/or

lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft

capsules, the active compounds may be dissolved or suspended in suitable liquids,

such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition,

5   stabilizers may be added.

       The above methodologies may be employed either actively or prophylactically

against an infection of interest.

Computer-related Aspects and Embodiments

       In addition to the provision of compounds as chemical entities, nucleotide

10   sequences, or fragments thereof at least 95%, preferably at least 97%, more preferably

at least 99%, and most preferably at least 99.9% identical to phage inhibitor sequences

can also be provided in a variety of additional media to facilitate various uses.

       Thus, as used in this section, "provided" refers to an article of manufacture,

rather than an actual nucleic acid molecule, which contains a nucleotide sequence of

15   the present invention; e.g., a nucleotide sequence of an exemplary bacteriophage or a

sequence encoding a bacterial target or a fragment thereof, preferably a nucleotide

sequence at least 95%, more preferably at least 99% and most preferably at least

99.9% identical to such a bacteriophage or bacterial sequence, for example, to a

polynucleotide of an unsequenced phage listed in Table 1, preferably of bacteriophage

20   77 (S. aureus host) or bacteriophage 3A (S.aureus host) or bacteriophage 96 (S.

aureus host). Such an article provides a large portion of the particular bacteriophage

genome or bacterial gene and parts thereof (e.g., a bacteriophage open reading frame

(ORF)) in a form which allows a skilled artisan to examine and/or analyze the

sequence using means not directly applicable to examining the actual genome or gene

25   or subset thereof as it exists in nature or in purified form as a chemical entity.

       In one application of this aspect, a nucleotide sequence of the present

invention can be recorded on computer readable media. As used herein, "computer

readable media" refers to any medium that can be read and accessed directly by a

computer. Such media include, but are not limited to: magnetic storage media, such

30   as floppy discs, hard disc storage medium, magnetic tape; optical storage media such

as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these

categories, such as magnetic/optical storage media. A skilled artisan can readily

appreciate how any of the presently known computer readable mediums can be used

to create an article of manufacture which includes one or more computer readable

35   media having recorded thereon a nucleotide sequence or sequences of the present

invention. Likewise, it will be clear to those of skill how additional computer

readable media that may be developed also can be used to create analogous manufactures having recorded thereon a nucleotide sequence of the present invention.

As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently 5 known methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence 10 of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can, for example, be presented in a word processing test file, formatted in 15 commercially available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data processor structuring formats (*e.g.*, text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence 20 information of the present invention.

Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. Thus, by providing in computer readable form a nucleotide sequence of an unsequenced bacteriophage, such as an exemplary bacteriophage listed in Table 1 or of a sequence 25 encoding a bacterial target or a fragment thereof, preferably a nucleotide sequence at least 95%, more preferably at least 99% and most preferably at least 99.9% identical to such a bacteriophage or bacterial sequence, for example, to a polynucleotide of bacteriophage 77 (*S. aureus* host) or bacteriophage 3A (*S.aureus* host) or bacteriophage 96 (*S. aureus* host), the present invention enables the skilled artisan to 30 routinely access the provided sequence information for a wide variety of purposes.

Those skilled in the art understand that software can implement a variety of different search or analysis software which implement sequence search and analysis algorithms, *e.g.*, the BLAST (Altschul et al., J. Mol. Biol. 215:403410 (1990) and BLAZE (Brutlag et al., Comp. Chem 17:203-207 (1993)) search algorithms. For 35 example, such search algorithms can be implemented on a Sybase system and used to identify open reading frames (ORFs) within the bacteriophage genome which contain homology to ORFs or proteins from other viruses, e.g, other bacteriophage, and other organisms, *e.g.*, the host bacterium. Among the ORFs discussed herein are protein

encoding fragments of the bacteriophage genomes which encode bacteria-inhibiting proteins or fragments.

The present invention further provides systems, particularly computer-based systems, which contain the sequence information described. Such systems are designed to identify, among other things, useful fragments of the bacteriophage genomes.

As used herein, "a computer-based system" refers to the hardware, software, and data storage media used to analyze the nucleotide sequence information of the present invention. The minimum hardware of the computer-based systems of the present invention comprises a central processing unit (CPU), input device, output device, and data storage medium or media. A skilled artisan will readily recognize that any of the currently available general purpose computer-based system are suitable for use in the present invention, as well as a variety of different specialized or dedicated computer-based systems.

As stated above, the computer-based systems of the present invention comprise data storage media having stored therein a nucleotide sequence of the present invention and the necessary hardware and software for supporting and implementing a search and/or analysis program.

As used herein, "data storage media" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present invention.

As used herein, "search program" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of the present gnomic sequences which match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBIA). A skilled artisan can readily recognize that any one of the available algorithms or implementing software packages for conducting homology searches and/or sequence analyses can be adapted for use in the present computer-based systems.

As used herein in connection with sequence searches and analyses, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in

SD-63311.4

the database. Also, the target sequence length is preferably selected to include sequence corresponding to a biologically relevant portion of an encoded product, for example a region which is expected to be conserved across a range of source organisms. Preferably the sequence length of a target polypeptide sequence is from 5-100 amino acids, more preferably 7-50 or 7-100 amino acids, and still more preferably 10-80 or 10-100 amino acids. Preferably the sequence length of a target polynucleotide sequence is from 15-300 nucleotide residues, more preferably from 21-240 or 21-300, and still more preferably 30-150 or 30-300 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length. Likewise, it may be desirable to search and/or analyze longer sequences.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzymatic active sites and signal sequences. Nucleic acid target motifs include, but are. not limited to promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

A variety of structural formats for the input and output devices can be used to input and output the information in the computer-based systems of the present invention. A preferred format for an output device ranks fragments of the bacteriophage or bacterial sequences possessing varying degrees of homology to the target sequence or target motif. Such presentation provides a skilled artisan with a ranking of sequences which contain various amounts of the target sequence or target motif and identifies the degree of homology contained in the identified fragment.

A variety of comparing methods and/or devices and/or formats can be used to compare a target sequence or target motif with the sequence stored in data storage media to identify sequence fragments of the bacteriophage or bacterium in question. One skilled in the art can readily recognize that any one of the publicly available homology search programs can be used as the search program for the computer-based systems of the present invention. Of course, suitable proprietary systems that may be known to those of skill, or later developed, also may be employed in this regard.

Figure 5 provides a block diagram of a computer system illustrative of embodiments of this aspect of present invention. The computer system 102 includes a processor 106 connected to a bus 104. Also connected to the bus 104 are a main memory 108 (preferably implemented as random access memory, RAM) and a variety

SD-63311.4

of secondary storage devices 110, such as a hard drive 112 and a removable medium storage device 114. The removable medium storage device114 may represent, for example, a floppy disk drive, a CD-ROM drive, a magnetic tape drive, etc. A removable storage medium 116 (such as a floppy disk, a compact disk, a magnetic

5   tape, etc.) containing control logic and/or data recorded therein may be inserted into the removable medium storage device 114. The computer system 102 includes appropriate software for reading the control logic and/or the data from the removable medium storage device 114, once it is inserted into the removable medium storage device 114.

10   A nucleotide sequence of the present invention may be stored in a well-known manner in the main memory 108, any of the secondary storage devices 110, and/or a removable storage medium 116. During execution, software for accessing and processing the sequence (such as search tools, comparing tools, etc.) reside in main memory 108, in accordance with the requirements and operating parameters of the

15   operating system, the hardware system and the software program or programs.

The data storage medium in which the sequence is embodied and the central processor need not be part of a single stand-alone computer, but may be separated so long as data transfer can occur. For example, the processor or processors being utilized for a search or analysis can be part of one general purpose computer, and the

20   data storage medium can be part of a second general purpose computer connected to a network, or the data storage medium can be part of a network server. As another example the data storage medium can be part of a computer system or network accessible over telephone lines or other remote connection method.

EXAMPLES

Example 1: Propagation of Bacteriophage 77 of *Staphylococcus aureus*

5   **Bacterial propagating strain and Bacteriophage:**

The Staphylococcus aureus propagating strain 77 (PS 77) was used as a host to propagate its respective phage 77 (ATCC # 27699-B1).

Purification of bacteriophage and prepration of phage DNA:

The propagation method was carried out by using the agar layer method

10   described by Swanstörm and Adams (Swanström, M. and Adams, M.H. (1951). Agar layer method for production of high titer phage stocks. *Proc. Soc. Exptl. Biol. & Med.* **78:** 372-375). Briefly, the PS 77 strain was grown overnight at 37°C in Nutrient broth [NB: 3 g Bacto Beef Extract, 5 g Bacto Peptone per liter, (Difco Laboratories)]. The culture was then diluted 20x in NB and incubated at 37°C until the $OD_{540}= .2$. The

15   suspension (15×10^7 Bacteria) was then mixed with 15×10^5 phage particles to give a ratio of 100 bacteria/phage particle in the presence of 400 μg/ml of $CaCl_2$. After incubation of 15 min at room temperature, 7.5 ml of melted soft agar (NB supplemented with 0.6% of agar), were added to the mixture and poured onto the surface of 100 mm nutrient agar plates (3 g Bacto Beef Extract, 5 g Bacto Peptone and

20   15 g of Bacto Agar per liter) and incubated overnight at 30°C. To collect the lysate, 20 ml of NB were added to each plate and the soft agar layer was collected by scrapping off with a clean microscope slide and shaken vigorously for 5 min to break up the agar. The mixture was then centrifuged for 10 min at 4,000 rpm and the supernatent (lysate) is collected and subjected to a treatment with 10 μg /ml of DNase I and

25   RNase A for 30 min at 37°C. To precipitate the phages particles, 10% (w/v) of PEG 8000 and 0.5 M of NaCl were added to the lysate and the mixture was incubated on ice for 16 h. The phages were recovered by centrifugation at 4,000 rpm for 20 min at 4°C on a GS-6R table top centrifuge (Beckman) . The pellet was resuspended with 2 ml of phage buffer (1 mM $MgSO_4$, 5 mM $MgCl_2$, 80 mM NaCl and .1% Gelatin). The

30   phage suspension was extracted with 1 volume of chloroform and purified by centrifugation using a TLS 55 rotor and the Optima TLX ultracentrifuge (Beckman), for 2 h at 28,000 Rpm at 4°C in preformed cesium chloride gradient as described in Sambrook et al. (Sambrook, J., Fritsch, E.F. and Maniatis, T (1989). Molecular

cloning: A laboratory Manual. Cold Spring Harbor Laboratory, New York. Cold Spring Harbor Laboratory Press). Banded phages were collected and ultracentrifuged again on an isopycnic cesium chloride gradient at 40,000 rpm for 24 h rpm at 4°C using a TLV rotor (Beckman). The phage was dialyzed for 4 h at room temperature against 4 L of dialysis buffer consisting of 10 mM NaCl, 50 mM Tris-HCl pH 8 and 10 mM $MgCl_2$. Phage DNA was prepared from the phages by adding 20 mM EDTA, 50 mg/ml Proteinase K and 0.5% SDS and incubating for 1 h at 65°C, followed by successive extractions with 1 volume of phenol, 1 volume of phenol-chloroform and 1 volume of chloroform. The DNA was then dialyzed overnight at 4°C against 4 L of T.E (10 mM Tris $_{8.0}$, 1mM EDTA).

Example 2: Preparation of Bacteriophage 77 DNA for Sequencing

Sonication of DNA:

4 µg of phage DNA was diluted in 200 µl of T.E pH 8.0 in a 1.5 ml Eppendorf tube and sonication was performed (550 Sonic Dismembrator, Fisher Scientific). Samples were sonicated under an amplitude of 3 µm with bursts of 5 s spaced by 15 s cooling in ice/water for 3 to 4 cycles and size-fractioned on 1% agarose gels. Fractions ranging from 1 to 2 kbp were isolated and gel purified by using the Qiagen kit according to the instructions of the manufacturer (Qiagen) and eluted in 50 µl of Tris 1mM, pH 8.5.

Repair of fragmented DNA ends:

The ends of the sonicated DNA fragments were repaired with a combination of T4 DNA polymerase and Klenow as follows. Reactions were performed in a final volume of 100 µl containing DNA, 10 mM Tris-HCl pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT, 5 µg BSA, 100 µM of each dNTP and 15 units of T4 DNA polymerase (New England Biolabs) for 20 min at 12°C followed by addition of 12.5 units of Klenow large fragment (New England Biolabs) for 15 min at room temperature. The reaction was stopped by two phenol/chloroform extractions and the DNA was ethanol precipitated and resuspended in 20 µl of H_2O.

Cloning into pKSII and transformation:

Blunt-ended DNA fragments were cloned by ligation directly into *Hin*II (New England Biolabs) and calf intestinal phosphatase (New England Biolabs)-treated pKSII vector (Stratagene). A typical reaction contained 100 ng of vector, 2 to 5 μl of repaired sonicated phage DNA in a final volume of 20 μl containing, 800 units of T4

5   DNA ligase (New England Biolabs) for overnight at 16 °C. Transformation and selection of positive clones was performed in the host strain DH10 β of E.coli using ampicillin as a selective antibiotic as described in Sambrook et al. (*supra*)

Preparation of sequencing templates:

Recombinant clones were picked from agar plates into 96-well plates. The

10 presence of foreign insert was confirmed by PCR analysis using T3 and T7 primers. PCR amplification of foreign insert was performed in a 15-μl reaction volume containing 10 mM Tris (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 0.02% gelatin, 1 μM primer, 187.5 μM each dNTP, and 0.75 units *Taq* polymerase (BRL). The thermocycling parameters were as follows: 2 min initial denaturation at 94°C for 2

15 min, followed by 20 cycles of 30 sec denaturation at 94°C, 30 sec annealing at 58°C, and 2 min extension at 72°C, followed by a single extension step at 72°C for 10 min. Clones with insert sizes of 1 to 2 kbp were selected and miniprep DNA of the selected clones were prepared using QIAprep spin miniprep kit (Qiagen).

20 Example 3: DNA Sequencing

DNA sequencing:

The ends of each recombinant clone were sequenced on an ABI 377-36 automated sequencer with two types of chemistry: ABI prism bigdye primer or ABI

25 prism bigdye terminator cycle sequencing ready reaction kit (Applied Biosystems). To ensure co-linearity of the sequence data and the genome, all regions of phage genome were sequenced at least once from both directions on two separate clones. In areas that this criteria was not met, a sequencing primer was selected and phage DNA was used directly as sequencing template employing ABI prism bigdye terminator

30 cycle sequencing ready reaction kit.

Sequence contig assembly:

Sequence contigs were assembled using Sequencher 3.1 software (GeneCodes). To close contig gaps, sequencing primers were selected near the edge of the contigs. Phage DNA was used directly as sequencing template employing ABI prism bigdye terminator cycle sequencing ready reaction kit.

5   The sequence obtained for phage 77 is shown in Table 2. The sequences for phage 3A and 96 were obtained by similar sequencing methods; the sequences of those phage genomes are shown in Tables 7 & 9 respectively.

Example 4: Sequence Analysis

10

Sequence analysis:

An implementation of the publicly available program SEQUIN, available for download at ftp://negi.nlm.nih.gov/sequin/, was used on phage genome sequence to identify all putative ORFs larger than 33 codons. A listing of such ORFs for *S.*

15   *aureus* phage 77 is shown in Table 3, with predicted amino acid sequences for selected ORFs shown in Table 4. Listings of ORFs for phage 3A and 96 are provided in Tables 8 and 10 respectively. A variety of other ORF identification could be used as alternatives and are known to those skilled in the art. Sequence homology searches for each ORF are then carried out using a standard implementation of blast programs.

20   Downloaded public databases used for sequence analysis include:
non-redundant GenBank (ftp://ncbi.nlm.nih.gov/blast/db/nr.Z),
Swissprot (ftp://ncbi.nlm.nih.gov/blast/db/swissprot.Z);
vector (ftp://ncbi.nlm.nih.gov/blast/db/vector.Z);
pdbaa databases (ftp://ncbi.nlm.nih.gov/blast/db/pdbaa.Z);

25   staphylococcus aureus NCTC 8325 (ftp://ftp.genome.ou.edu/pub/staph/staph-1k.fa);
streptococcus pyogenes (ftp://ftp.genome.ou.edu/pub/strep/strep-1k.fa);
streptococcus pneumoniae
(ftp://ftp.tigr.org/pub/data/s_pneumoniae/gsp.contigs.112197.Z);
mycobacterium tuberculosis CSU#9

30   (ftp://ftp.tigr.org/pub/data/m_tuberculosis/TB_091097.Z); and
pseudomonas aeruginosa (http://www.genome.washington.edu/pseudo/data.html).

Exemplary results of homology searches are shown in Table 5 for bacteriophage 77.

Example 5: Identification of Cecropin Signature Motif in *Staphylococcus aureus* Bacteriophage 3A ORF

The genome for *S. aureus* bacteriophage 3A was determined and the sequence was analyzed essentially as described for bacteriophage 77 in the examples above. Upon blast analysis of the identified open reading frames of phage 3A, the presence of an amino acid sequence corresponding to a cecropin signature motif was observed. This motif (WDGHKTLEK) is located at position aa 481-489. Cecropins were originally identified in proteins from the cecropia moth and are recognized as potent antibacterial proteins that constitute an important part of the cell-free immunity of insects. Cecropins are small proteins (31-39 amino acid residues) that are active against both Gram-positive and Gram-negative bacteria by disrupting the bacterial membranes. Although the mechanisms by which the cecropons cause cell death are not fully understood, it is generally thought to involve channel formation and membrane destabilization.

The identification of a motif corresponding to a known inhibitor suggests that the product of ORF002 is also an inhibitory compound. Such inhibitory activity can be confirmed as described herein or by other methods known in the art. Confirmation of the inhibitory activity would indicate that the ORF product could serve as the basis for construction of mimetic compounds and other inhibitors directed to the target of the ORF002 product.

Boman & Hultmark, 1987, *Ann. Rev. Microbiol.* 41:103-126.

Boman, 1991, *Cell* 65:205-207.

Boman et al., 1991, *Eur. J. Bioichem.* 201:23-31.

Wang et al., *J. Biol. Chem.* 273:27438-27448.

Example 6: Bacteriophage 77 ORF Expression

Bacteriophage ORFs are prepared and expressed as generally described in the Detailed Description above, utilizing a shuttle expression vector with a locus for insertion of a phage ORF subject to inducible expression in an appropriate host bacterium.

Preparation of shuttle expression vector:

The shuttle vector pT0021, in which the firefly luciferase (*lucFF*) expression is controlled by the *ars* promoter/operator from a *S. aureus* plasmid (Tauriainen, S., Karp, M., Chang, W and Virta, M. (1997). Recombinant luminescent bacteria for measuring bioavailable arsenite and antimonite. *Appl. Environ. Microbiol.* **63**:4456-4461), was modified as below to suit our specific application. Two oligonucleotides corresponding to the influenza HA tag were synthesized. The sense strand HA tag sequence (with *BamH*I, *Sal*I and *Hind*III cloning sites) is:

5'-gatcccggtcgaccaagcttTACCCATACGACGTCCCAGACTACGCCAGCTGA-3';

the antisense strand HA tag sequence (with *Hind*III cloning site) is:

5'-agctTCAGCTGGCGTAGTCTGGGACGTCGTATGGGTAaagcttggtcgaccgg-3'.

The two HA tag oligonucleotides were annealed following a standard protocol (*supra*) and ligated to pT0021 vector that was digested with *BamH*I and *Hind*III (the *lucFF* gene was released from the vector and replaced by the HA tag). This modified shuttle vector containing the *ars* promoter, *arsR* gene and HA tag was named pTHA vector.

Cloning of ORFs with a Shine-Dalgarno sequence:

ORFs with a Shine-Dalgarno sequence were selected for functional analysis of bacterial killing. Each ORF, from initiation codon to last codon (excluding the stop codon), was PCR amplified from phage genomic DNA. For PCR amplification of ORFs, each sense strand primer starts at the initiation codon and is preceded by a *BamH*I restriction site and each antisense strand starts at the last codon (excluding the stop codon) and is preceded by a *Sal* I restriction site. PCR product of each ORF was gel purified and digested with *BamH*I and *Sal*I overnight. The digested PCR product was then gel purified, ligated into *BamH*I and *Sal*I digested pTHA vector, and used to transform bacterial strain DH10β. As a result, HA tag is inframe with the ORF and a fusion protein with ORF begins at N-terminal and HA tag ends at the C-terminal is produced. Recombinant ORF clones were picked and their sizes were confirmed by PCR analysis using primers flanking the cloning site. The sequence fidelity of cloned ORFs was verified by DNA sequencing using the same primers as used for PCR. In the cases that the verification of ORFs could not be achieved by one path of sequencing using primers flanking the cloning site, internal primers were selected and used for sequencing.

Transformation of *Staphylococcus aureus* with expression constructs

Staphylococcus aureus strain RN4220 (Kreiswirth et al., 1983, Nature 305:709-712) was used as a recipient for the expression of recombinant plasmids. Electroporation was performed essentially as previously described (Schenk and Laddaga, 1992, FEMS Microbiology Letters 94:133-138). Selection of recombinant

5   clones was performed on Luria-Broth agar (LB-agar) plates containing 30 µg/ml of Kanamycin.

Chemical inducers

Sodium arsenite ($NaAsO_2$), sodium arsenate (Na_2HAsO_4), and antimony potassium tartrate ($K(SbO)C_4H_4O_6$) were purchased from Sigma (Sigma-Aldrich

10   Canada LTD, Oakville) and were used as heavy metals to induce gene expression from the ars promoter/operator.

Induction of gene expression from the ars operon

Cells containing different recombinant plasmids were grown overnight at 37°C in LB medium supplemented with 30 µg/ml of Kanamycin. The cells were then

15   diluted to the mid log phase (OD_{540} approx. 0.2) with fresh LB media containing Kanamycin and transferred to 96-well microtitration plates (100 µl/well). Inducers were then added at different final concentrations (ranging from 2.5 to 10 µM) and the culture was incubated for an additional 2 h at 37°C. Control cultures without inducers were cultured in separate wells. The effect of expression of the phage 77 ORFs on

20   bacterial cell growth was then monitored by measuring the OD540 and comparing the rate of growth of the culture containing inducer to the rate of growth of the culture not containing inducer. As positive controls for growth inhibition, the kilA gene of phage lambda (Reisinger et al., 1993, Virology 193:1033-1036), and the holin/lsinI genes of the Staphylococcus aureus phage Twort (Loessner et al., 1998, FEMS Microbiology

25   Letters 162:265-274) were subcloned into the ars inducible vector and included in separate wells of the microtitration plate.

Expression of ORFs from a large variety of other phage can be accomplished using the above vector, or other vector adapted for an appropriate bacterium and preferably for inducible expression of the insert ORF or ORFs.

30

All patents and publications mentioned in the specification are indicative of the levels of skill of those skilled in the art to which the invention pertains. All

references cited in this disclosure are incorporated by reference to the same extent as if each reference had been incorporated by reference in its entirety individually.

One skilled in the art would readily appreciate that the present invention is well adapted to carry out the objects and obtain the ends and advantages mentioned, as well as those inherent therein. The specific methods and compositions described herein as presently representative of preferred embodiments are exemplary and are not intended as limitations on the scope of the invention. Changes therein and other uses will occur to those skilled in the art which are encompassed within the spirit of the invention are defined by the scope of the claims.

It will be readily apparent to one skilled in the art that varying substitutions and modifications may be made to the invention disclosed herein without departing from the scope and spirit of the invention. For example, those skilled in the art will recognize that the invention may suitably be practiced using a variety of different bacteria, bacteriophage, and sequencing methods within the general descriptions provided.

The invention illustratively described herein suitably may be practiced in the absence of any element or elements, limitation or limitations which is not specifically disclosed herein. Thus, for example, in each instance herein any of the terms "comprising," "consisting essentially of" and "consisting of" may be replaced with either of the other two terms. The terms and expressions which have been employed are used as terms of description and not of limitation, and there is not intention that in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention claimed. Thus, it should be understood that although the present invention has been specifically disclosed by preferred embodiments and optional features, modification and variation of the concepts herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this invention as defined by the appended claims.

In addition, where features or aspects of the invention are described in terms of Markush groups or other grouping of alternatives, those skilled in the art will recognize that the invention is also thereby described in terms of any individual member or subgroup of members of the Markush group or other group. For example, if there are alternatives A, B, and C, all of the following possibilities are included: A separately, B separately, C separately, A and B, A and C, B and C, and A and B and C. Thus, for example, for the bacteria and phage specified herein, the embodiments expressly include any subset or subgroup of those bacteria and/or phage. While each

SD-63311.4

such subset or subgroup could be listed separately, for the sake of brevity, such a listing is replaced by the present description.

Thus, additional embodiments are within the scope of the invention and within the following claims.